Grid Alignment in Entorhinal Cortex

Bailu Si,[1] Emilio Kropff,[2] and Alessandro Treves[1, 2]

[1]*Sector of Cognitive Neuroscience, International School for Advanced Studies, via Bonomea 265, 34136 Trieste, Italy**

[2]*Kavli Institute for Systems Neuroscience and Center for the Biology of Memory,*
Norwegian University of Science and Technology, 7489 Trondheim, Norway†

The spatial responses of many of the cells recorded in all layers of rodent medial entorhinal cortex (mEC) show a triangular grid pattern, which appears to provide an accurate population code for position, and once established might be based in part on path-integration mechanisms. Competing models, each partially contradicted by experimental observations, try to explain how the grid-like pattern emerges in terms of network interactions, or of interactions with theta oscillations or, the one we have proposed, of mere single-unit mechanisms.

Grid axes are tightly aligned across simultaneously recorded units. Recent experimental findings have shown that grids can often be better described as elliptical rather than purely circular and that, beyond the mutual alignment of their grid axes, ellipses tend to also orient their long axis along preferred directions. Are grid alignment and ellipse orientation the same phenomenon? Does the grid alignment result from single-unit mechanisms or does it require network interactions?

We address these issues by refining our model, to describe specifically the spontaneous emergence of conjunctive grid-by-head-direction cells in layers III, V and VI of mEC. We find that tight alignment can be produced by recurrent collateral interactions, but this requires head-direction modulation. Through a competitive learning process driven by spatial inputs, grid fields then form already aligned, and with randomly distributed spatial phases. In addition, we find that the self-organization process is influenced by the behavior of the simulated rat. The common grid alignment often orients along preferred running directions. The shape of individual grids is distorted towards an ellipsoid arrangement when some speed anisotropy is present in exploration behavior. Speed anisotropy on its own also tends to align grids, even without collaterals, but the alignment is seen to be loose. Finally, the alignment of spatial grid fields in multiple environments shows that the network expresses the same set of grid fields across environments, modulo a coherent rotation and translation. Thus, an efficient metric encoding of space may emerge through spontaneous pattern formation at the single-unit level, but it is coherent, hence context-invariant, if aided by collateral interactions.

Keywords: Hippocampus; Entorhinal cortex; Grid cells; Conjunctive grid-by-head-direction cells; Firing rate adaptation; Competitive network; Remapping

I. INTRODUCTION

Internal representations of space appear necessary for any agent, such as a rat or a robot, to move around in a spatial context and to distinguish between different contexts to which food or danger, for example, may be associated. Spatial cognition and memory have been long investigated in rodents, and an impressive body of results point at the major role being played by the hippocampus and related cortices, a region that in humans has been associated with the neural basis of episodic memory formation, since[1]. Forty years ago, place cells were discovered in the rat hippocampus, showing specific firing activity whenever the rat enters a specific portion of the environment, the place field[2]. Head direction (HD) cells were first found in the rat postsubiculum, firing steadily when the animal points its head towards a specific direction in the environment[3]. These two distinct systems provide simple, distributed population coding of the location and heading direction of the animal.

In recent years, cells with more complex spatial codes have been discovered in the rat medial entorhinal cortex (mEC), which sends strong projections to the hippocampus. Grid cells, found to be particularly abundant in layer II of mEC (perhaps about half of stellate cells there)

have multiple firing fields positioned on the vertices of remarkably regular triangular grids, spanning the environment which the animal explores[4,5]. Conjunctive grid-by-head-direction cells, found along a smaller proportion of pure grid cells in the deeper layers of mEC, show firing selectivity to HD in addition to (perhaps slightly less precise) spatial tuning as grid cells[6]. The precise geometric tessellation provided by the activity of grid cells has stimulated a series of experimental and theoretical studies on the mechanisms underlying the emergence and the function of grid cells[7].

Theoretical models of grid cell formation may be grouped in three main categories. The first type of models shows how grid fields may emerge from the attractor states induced, in continuous attractor networks, from structured recurrent connectivity[8–12]. The spatial layout of recurrent collateral connections of the network, assumed to be permanent or at least present during the developmental stage of grid cell formation, ensures that triangular spatial firing patterns are stable states of recurrent dynamics. Grid units in a continuous attractor network model are able to perform path integration by propagating the activity in the network in correspondence with the movement of the animal, in a way similar to previous models for place cells or HD cells[13–15].

The second class of models relates the periodic firing of grid cells to sub-threshold membrane potential oscillations, and proposes that grid fields may result from interference between a theta-related baseline oscillator and other velocity-controlled oscillations which originate either in different dendrites of a neuron or in different neurons[16–20]. The frequency difference between these velocity-controlled oscillators is small, so that the low frequency "envelope" of the interference pattern corresponds to the spatial periodicity of grid cells. This hypothesis is in part supported by recent findings that the spatial periodicity of grid cells is susceptible to the suppression of theta oscillations by pharmacological silencing of the medial septum[21,22].

The third class of models on grid cell formation argues that grid fields may not require detailed ad hoc mechanisms like structured connectivity or theta oscillations, rather they may emerge spontaneously from a general feature of cortical cell activity, like firing rate adaptation[23] or, equivalently, other types of temporal modulation[24]. Such temporal modulation is shown in computer simulations to sculpt the spatial modulation of grid cells through a self-organization process that, averaged over a long developmental time of one or two weeks[25,26], leaves as a footprint on each unit the regular periodicity found in real grid cells. A simple analytical model "explains" this spontaneous pattern formation as an unsupervised optimization process at the single-unit level[23].

An interesting aspect of this model, which motivates the present study, is that the emergence of perfect grid symmetry requires a perfectly isotropic distribution of rat trajectories and speed, once averaged over the long (but not infinite) learning period. Any deviation from perfect isotropy, for example because the animal spends the relevant developmental period (for a rat, somewhere between P15-P35, say) mainly in a rectangular cage and tends to move along the walls, or because it runs a bit faster along some particular directions, would be expected to induce distortions in all grids units. Excitingly, such deviations in the geometry of grid maps have been observed recently and described as an *ellipticity* effect[27]. This phenomenon cannot be explained as small random deviations from the perfect symmetry, because of its extent and remarkable consistency across the population. Not only, as observed earlier, do grid axes show nearly the same alignment across all simultaneously recorded grid units, but also the long axes of the corresponding ellipses appear to loosely orient with either one of two major ellipse clusters, depending on their spacings[27].

Can the mutual alignment result from the same single-unit mechanisms that produce the individual grid patterns? With perfect isotropic grids, the answer is obviously negative, and in fact Kropff and Treves (2008) pointed at a mechanism that can align grid cells at a population level, but based on collateral interactions between them. With the observed anisotropies, however, the situation might be different, as both alignment and common orientation might develop along the (common) anisotropy axes.

In the computer simulation study reported here we find, again, that developing a tight alignment requires network interactions, through recurrent collateral connections with a specific structure, modulated by HD. As a first focus of this paper, we then describe how collateral connections may align grid cells, and then link this to the emergence of a coherent population *ellipticity* when the behavior of the animal is biased by running direction or speed. We then argue that anisotropy at the single-unit level is in principle sufficient to also produce some alignment, but fails in practice to make different units as tightly aligned as experimentally observed.

Finally, we also investigate how well the same model describes *global remapping*[28]. When a rat is taken into a new environment, or when the environment is manipulated in a way that it looks new, place maps in the hippocampus undergo an apparently random shuffling, or switch from active to inactive and vice versa. Grid maps, however, behave coherently at the population level, undergoing a common rotation and spatial shift, in such a way that spatial overlaps between maps are preserved across rooms[29]. We thus address the question of how, within our model, coherent grid fields may emerge in the same network for multiple environments.

The rest of the article is organized as follows. The network model is first introduced in Section II. In Section III grid alignment in a cylinder environment is studied, and it is compared to what occurs in a square environment in Section IV. We investigate in Section V the effects of exploration with speed anisotropy. Grid realignment in multiple environments is discussed in Section VI. Finally, the results of the study are discussed in Section VII.

II. NETWORK MODEL

In Fig. 1 we present a diagram of the network that we use for simulations. It is intended to model conjunctive grid-by-head-direction cells in layers III, V and VI of mEC. In layer V, which receives strong projections from the subiculum and the CA1 region of the hippocampus, about 20% of the putative pyramidal cells are estimated to be conjunctive grid-by-head-direction cells, along with a very small proportion of pure grid cells and more than 60% of HD cells[30]. In layer III, which projects to CA1, the proportion of conjunctive cells is similar to that of layer V, but the proportion of HD cells is much lower (20%) and there is an extra 20% of pure grid cells. These layers present a prominent recurrent connectivity, denser in layer V (around 12%) but also important in layer III (around 9%)[31]. Layer II of mEC, where the highest proportion and the best quality of pure grid cells are found (around 50%), lacks two critical elements that in our model go hand by hand: recurrent collateral connections and head direction information. In this paper, we assume that layer II maps can self-organize using inputs which include conjunctive maps, an assumption that is discussed
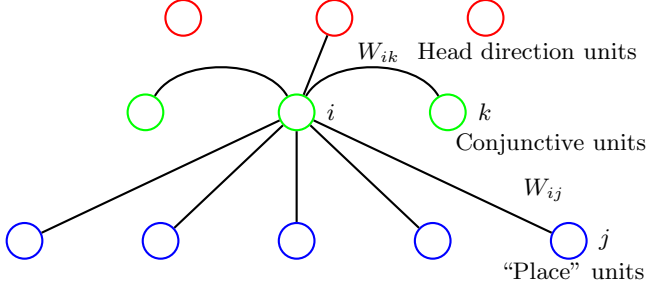
Figure 1: A sketch of the network model for conjunctive cells in deep layers of mEC. Connections are shown only for unit i. Place units are fully connected to conjunctive units. Conjunctive units are connected to all other conjunctive units without self-connections. Each conjunctive unit is assumed to be modulated by one HD unit, representing the overall effect of angular modulation from the local network.

in a separate study currently in progress. Here, we focus therefore on the learning process in layers III to VI, and neglect the possible influence there of the feedback from layer II.

The critical difference with previous versions of the model[23] is that now we assign a central role to head direction information, which, as we show, is a key element for grid alignment. A preferred head direction, θ_i, is introduced arbitrarily for each conjunctive unit i, to modulate its inputs. θ_i is uniformly sampled from all angles. This assumption is based on our own analysis of the HD selectivity of the conjunctive cells recorded by[6], in which we did not find significant clustering of HDs with respect to either the reference frame or grid axes. Each conjunctive unit receives afferent spatial inputs, which as discussed in[23] we take for simplicity to arise from regularly arranged "place" units, and collateral inputs from other conjunctive units (Fig. 1). The overall input to conjunctive unit i at time t is then given by h_i^t

$$h_i^t = f_{\theta_i}(\omega_t)(\sum_j W_{ij}^{t-1} r_j^t + \rho \sum_k W_{ik} \Psi_k^{t-\tau}), \quad (1)$$

with $\rho = 0.2$ a factor weighing, relative to feed-forward inputs, collateral inputs relayed with a delay $\tau = 25$ steps. In the model, each time step corresponds to 10 msec in real time, so the collateral interaction describe temporally diffuse and delayed processes, rather than straightforward AMPA-mediated excitation. $f_{\theta_i}(\omega_t)$ is a tuning function that has maximal value when the current HD ω^t of the simulated rat is along the preferred direction θ_i of the unit, as in[13]

$$f_\theta(\omega) = c + (1-c)\exp[\gamma(\cos(\theta - \omega) - 1)], \quad (2)$$

where $c = 0.2$ and $\gamma = 0.8$ are parameters determining the minimal value and the width of HD tuning.

The weight W_{ij}^t connects place unit j to conjunctive unit i, while W_{ik} connects conjunctive unit k to unit i. For simplicity, throughout the paper we only consider

the self-organization of the feed-forward weights, keeping the collateral weights fixed at convenient values, which we will discuss in Sect. II B. A model in which collateral weights are also the result of a perhaps slower self-organizing learning process will be discussed elsewhere.

It is possible that conjunctive cells receive place-cell-like inputs from the hippocampus already when they develop their firing maps. Studies on the development of the spatial representation system in the rat may be taken to show that place cells and head-direction cells develop adult-like spatial and directional codes somewhat earlier than grid cells do[25,26]. The firing rate of a place unit j is modeled by an exponential function centered in its preferred firing location $\vec{x}_{j0}$

$$r_j^t = \exp(-\frac{|\vec{x}^t - \vec{x}_{j0}|^2}{2\sigma_p^2}), \quad (3)$$

where $\vec{x}^t$ is the current location of the simulated rat. $\sigma_p = 5cm$ is the width of the firing field. The place field of a place unit is arranged such that the distances to neighboring fields of other place units are about $5cm$. Note that the precise and regularly arranged location-specific inputs of place units are used only to speed up the learning process. The network model works in qualitatively the same way with broad spatial inputs, but at the cost of longer learning time[23].

The firing rate Ψ_i^t of conjunctive unit i is determined by

$$\Psi_i^t = \Psi_{sat} \arctan[g^t(\alpha_i^t - \mu^t)]\Theta(\alpha_i^t - \mu^t), \quad (4)$$

where $\Psi_{sat} = \pi/2$, so that the maximal firing rate is 1 (in arbitrary units). $\Theta(\cdot)$ is the Heaviside function. The variable α_i^t represents a time-integration of the input h_i, adapted by the dynamical threshold β_i

$$\begin{aligned} \alpha_i^t &= \alpha_i^{t-1} + b_1(h_i^{t-1} - \beta_i^{t-1} - \alpha_i^{t-1}), \\ \beta_i^t &= \beta_i^{t-1} + b_2(h_i^{t-1} - \beta_i^{t-1}), \end{aligned} \quad (5)$$

where β_i has slower dynamics than α_i (b_2 is set to $b_2 = b_1/3$, and in our time steps corresponds to a time scale of roughly $300\ (ms)^{-1}$, with $b_1 = 0.1 \simeq 100(ms)^{-1}$). Due to the adaptation dynamics, a unit that fires strongly in the recent past tends not to fire in the near future, because of a high threshold.

In Eq. 4, the gain g^t and threshold μ^t are two parameters chosen at each time step to keep the mean activity $a = \sum_i \Psi_i^t / N_{mEC}$ of the conjunctive units and the sparsity $s = (\sum_i \Psi_i^t)^2 / (N_{mEC} \sum_i \Psi_i^{t\,2})$ within a 10% relative error bound from pre-specified values, set at $a_0 = 0.1$ and $s_0 = 0.3$ respectively. The values of g^t and μ^t are determined through iterations

$$\begin{aligned} \mu^{t,l+1} &= \mu^{t,l} + b_3(a^l - a_0), \\ g^{t,l+1} &= g^{t,l} + b_4 g^{t,l}(s^l - s_0). \end{aligned} \quad (6)$$

Here l is the index of the iteration within simulation step t. $b_3 = 0.01$ and $b_4 = 0.1$ are positive step sizes in iteration. a^l and s^l are the mean and the sparsity of the

conjunctive units when the gain $g^{t,l}$ and threshold $\mu^{t,l}$ are applied. The gain and threshold at the end of the iteration are used in Eq. 4 to determine the activity of conjunctive units.

With constant running speed, neural fatigue is invariant with respect to all directions. Conjunctive units self-organize firing fields into hexagonal/triangular grids in two-dimensional space, as the virtual rat explores the environment. This configuration of fields corresponds to the minimum of an energy function[23], consistent with an hexagonal tiling being the most compact one to arrange circles in two-dimensional space[32].

A. Learning feed-forward weights

The feed-forward weights are adaptively modified according to a Hebbian rule

$$\tilde{W}_{ij}^t = W_{ij}^{t-1} + \epsilon(\Psi_i^t r_j^t - \bar{\Psi}_i^{t-1} \bar{r}_j^{t-1}). \tag{7}$$

Here $\bar{\Psi}_i^t$ and $\bar{r}_j^t$ are estimated mean firing rates of conjunctive unit i and place unit j

$$\begin{aligned} \bar{\Psi}_i^t &= \bar{\Psi}_i^{t-1} + \eta \left(\Psi_i^t - \bar{\Psi}_i^{t-1} \right), \\ \bar{r}_j^t &= \bar{r}_j^{t-1} + \eta \left(r_j^t - \bar{r}_j^{t-1} \right), \end{aligned} \tag{8}$$

while $\epsilon = 0.005$ is a moderate learning rate, intended to produce gradual weight change, and $\eta = 0.05$ is a time averaging factor.

After each learning step, the new weights are normalized into unitary norm

$$\sum_j W_{ij}^t{}^2 = 1. \tag{9}$$

Through competitive learning, conjunctive units that win the competition are associated to the input units that provide strong inputs. As a result, firing fields of conjunctive units are anchored to places where inputs are strong simultaneously with recovery from adaptation, and are stabilized as the learning proceeds. There is experimental evidence of the dependence of grid fields on sensory information[33,34]. For example, grids can expand and contract in response to expansions and contractions of a familiar environment. Some still unpublished observations indicate that the learning of stable grid maps in a novel environment for non-over-trained adult rats could take several days of training[35].

B. Collateral weights

Two characteristic features of grid cells are that they align their axes and maintain fixed spatial phases relative to each other. Such alignment and relative phases are invariant to environmental changes, including those under which place cells in the hippocampus remap completely.
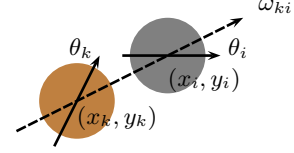

Figure 2: Assignment of the fixed collateral weight from unit k to unit i.

In our network model, we include these two constraints on grid cell population activity with the help of collateral weights. Head direction information is used to gate the interaction between two units that fire in sequence, so that the second unit will develop fields along its preferred direction close to those of the first unit, which fires earlier. If the interaction were not gated by HD, the second unit would tend to form a field that is a ring surrounding the field of the first unit.

The collateral weights are assigned before any learning takes place in the feed-forward weights. The weight W_{ik} for the connection from unit k to i is calculated in the following way. Each conjunctive unit i is temporarily assigned an auxiliary field, the location (x_i, y_i) of which is randomly chosen among the place fields in the environment. Once the weights are fixed, these auxiliary fields have nothing to do with the conjunctive units any longer. The collateral weight between unit k and i is calculated as

$$W_{ik} = [f_{\theta_k}(\omega_{ki}) f_{\theta_i}(\omega_{ki}) \exp(-\frac{d_{ki}^2}{2\sigma_f^2}) - \kappa]^+, \tag{10}$$

where $[\cdot]^+$ is a threshold function, with $[x]^+ = 0$ for $x < 0$, and $[x]^+ = x$ otherwise. $\kappa = 0.05$ is an inhibition parameter to favor sparse weights. $f_\theta(\omega)$ is the HD tuning function defined in Eq. 2. ω_{ki} is the direction of the line from field k to i. $\sigma_f = 10cm$ is the width of spatial tuning. $d_{ki} = \sqrt{[x_i - (x_k + \ell \cos \omega_{ki})]^2 + [y_i - (y_k + \ell \sin \omega_{ki})]^2}$ is the distance between field k and i with offset $\ell = 10cm$.

Normalization is performed on W_{ik} similarly as in Eq.9. The resulting weight structure allows strong collateral interactions between units that have similar HDs, and meanwhile avoids co-activation, to produce grids with certain spatial shifts.

In the following sections, we are going to show how, with the same set of fixed collateral weights and adaptable feed-forward weights, units in the network develop grid fields in single and multiple environments with various boundary shapes, and with different exploration behaviors of the virtual rat.

III. GRID ALIGNMENT IN CYLINDER ENVIRONMENTS

To test whether triangular grid fields can emerge and mutually align in our network model, we first simulate the

network for a virtual rat randomly exploring a cylinder environment, where trajectories are completely isotropic. For simplicity, we assume that when the rat runs in a certain direction, it points its head toward the running direction (RD) most of the time. Therefore, in our simulations the head direction and the running direction are not distinguished.

Each step in the simulation is taken to correspond to $10msec$ of real time. Each run of a simulation lasts for 8×10^6 steps, or over $22h$ real time. The standard speed of the rat is $v_s = 0.4m/s$, towards the peak speed of real rats in active exploration, so that, considering the time a real rat spends not in exploration, a run of the simulation is intended to correspond to the developmental time scale for the emergence of grid cells[25,26]. At each step, the running direction is chosen by using a pseudorandom Gaussian distribution with mean equal to the direction in the previous time step and an angular standard deviation $\sigma_{RD} = 0.2$ radians. Very importantly, the new direction cannot lead the rat outside the limits of the environment. If so, the selection process is repeated until a valid direction is chosen. This has the effect of favoring trajectories along the walls of the environment. The overall appearance of the trajectory is comparable with those of actual rats, reflecting also the natural tendency of naïve animals to run along the walls of the enclosure. The number of place units is 500 and the number of conjunctive units is 250. Place units are fully connected to conjunctive units. Each conjunctive unit receives connections from all other conjunctive units except itself. In the following simulations, we use the above mentioned default parameters, including the number of units in the network, the number of simulation steps, the speed and the noise in RD, unless stated differently.

In the first simulation covered in this section, the rat runs in a cylinder environment with a diameter of $125cm$. Fig. 3a plots part of a typical trajectory. The RD of the rat in this environment is uniformly distributed (Fig. 3b), making the cylinder environment a suitable control condition to compare to anisotropy later.

Fig. 4a shows four examples of the spatial maps of the conjunctive units in the network. The multiple firing fields of the units locate on the vertices of triangular grids. To better reveal grid structure, the autocorrelograms of the spatial firing maps are calculated (Fig. 4b). The {x,y} pixel in an autocorrelogram is the correlation between the original firing map and its shifted version by the vector {x,y}. The peaks away from the center of the autocorrelogram indicate the directions and distances at which the shifted map has a high overlap with the unshifted one. The six maxima that are closer to the center of each autocorrelogram are then detected, and the three maxima with positive y coordinates (white markers in Fig. 4b) define the orientations of the three grid axes (the autocorrelogram is redundant by definition, the lower part resulting from the rotation of the upper part by 180 degrees). The orientation of the first grid axis, i.e. the one with the lowest angle with respect to the x axis,

is used as the orientation of a grid. The mean distance of the three maxima away from the center is defined as the spacing of a grid. The average spacing of the grids that self-organize in our network is about 58cm, comparable to the grid spacings observed in experiments[5]. As noted previously, the spacing results from the parameters describing neuronal fatigue in the model[23].

Averaged into angular bins, the angular map of the firing of each unit shows HD selectivity, matching the HD tuning curve of the unit (Fig. 4c). This obviously results from the HD modulation on the inputs to conjunctive units (Eq. 1-2).

The spatial periodicity of conjunctive units is measured by the gridness score, as in[6]. We thus cut a ring area in each autocorrelogram with inner and outer radii chosen so as to contain the six maxima closest to the center and exclude the rest. We then correlate the original ring with its rotated versions, expecting to obtain for a perfect grid a very positive (negative) correlation coefficient for rotations at even (odd) multiples of 30 degrees. The gridness index is defined as the difference between the mean correlation for even rotations (60 and 120 degrees), and odd rotations (30, 90 and 150 degrees). Note that the gridness score is in the range [-2, 2].

The histogram of the gridness scores of all units is presented in Fig. 3c. Most units have high gridness score. Fig. 3d shows for all the 250 conjunctive units in the simulation the locations of the three maxima defining the grid axes. They clearly cluster into three clouds, indicating similar orientations and spacings for all units. The orientation histograms of the three grid axes are displayed in Fig. 3e, one color for each. To quantify the coherence in aligning grids, we define a *cross-population alignment coherence score* by the standard deviation of the orientations averaged over the three grid axes, with low values indicating tight grid alignment. In the simulation shown in Fig.3, the alignment coherence score is 2.967 degrees, indicating (very) tight grid alignment.

While grids are aligned, the spatial phases of all grids (shown relative to the best grid, as a conventional reference) are distributed in an area with diameter similar to the grid spacing (Fig. 3f). This means that grids cover the whole environment more or less randomly.

The collateral weights are sparse and are stronger for units with similar HDs (broken lines in Fig. 4d), a relationship given by Eq. 10. When the connectivity of collateral connections is reduced to 50%, the network is still able to align grids (data not shown). Therefore, the exact strength of the collateral connectivity is not critical for the model. Collateral weights are strongest between units that are active in locations a small distance apart and that are moderately correlated in firing, due to the delay parameter τ in their inputs (Fig. 4e). The network is not too sensitive to τ as long as $\tau \geq 13$ steps, i.e. $130msec$ in real time. This is the time it takes the virtual rat to travel the minimal distance between two input place fields. Smaller delays do not change the gridness scores or the alignment of grid maps, but cause the
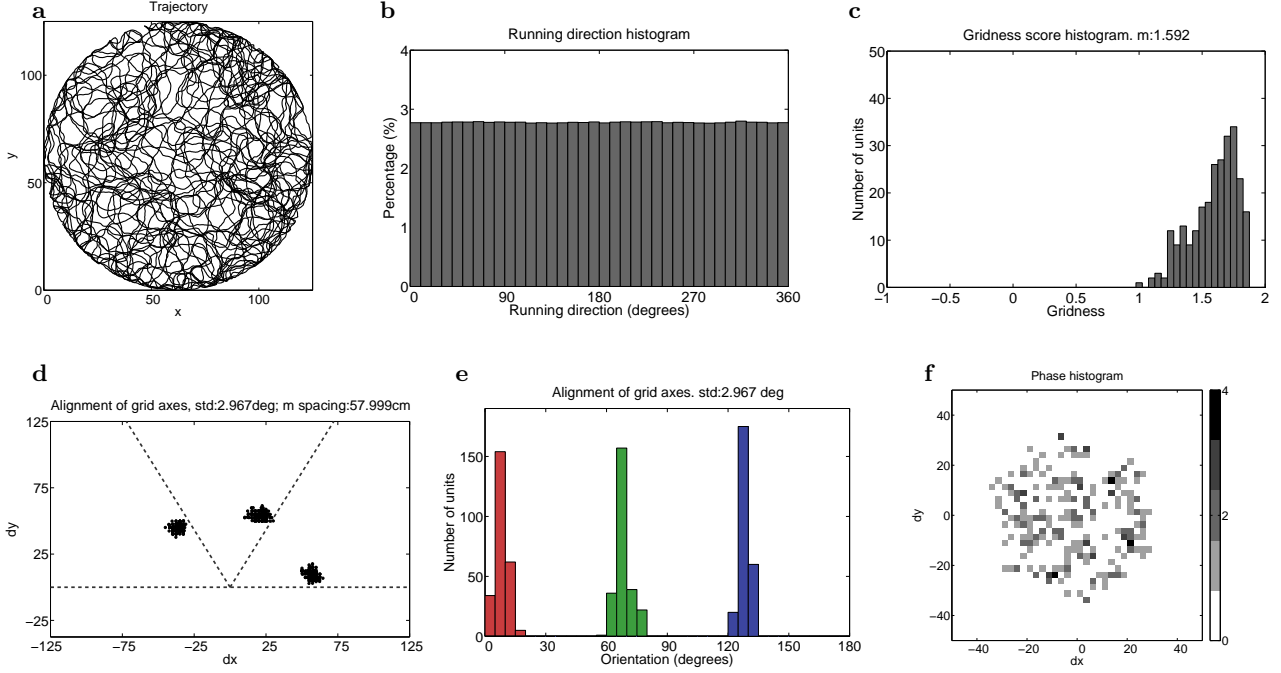
Figure 3: Grid alignment in a cylinder environment. (a) A trajectory of the virtual rat with 3×10^4 steps; (b) RD distribution of the trajectory in the simulation; (c) Histogram of the gridness scores of all conjunctive units in the simulation; (d) Scatter plot of the locations of the three peaks found in autocorrelograms (as shown by the white markers in Fig. 4b); (e) Histograms of the angles of the three grid axes, plotted with different colors for each axis. The alignment coherence score, i.e. standard deviation (in degrees) of the orientation averaged over the three grid axes, is indicated at the top of the panel; (f) Two dimensional histogram of spatial phases relative to the best grid. The grayscale encodes the number of units, the phases of which fall into a $2.5cm \times 2.5cm$ spatial bin. White color represents zero, and darker colors represent larger numbers.

spatial phases of the grid fields to collapse, showing a non-random phase distribution (data not shown). The phase collapse due to small τ can be understood by the associative learning between conjunctive units and place units. Small τ would cause two strongly connected conjunctive units to fire consecutively in short intervals, during which the place units show similar population activity. Associative learning therefore would produce similar grid fields for both units. To avoid that, the delay between two strongly connected conjunctive units needs to be long enough so that grid fields are able to anchor to different locations, resulting in random spatial phases. The exact mechanism that generates this delay is something that our simple rate-based model does not attempt to describe, and it could go from several types of slow synaptic potentials to interactions with rhythms, such as phase precession.

In summary, with fixed collateral weights, the grid fields of model units align to a common orientation while the relative spatial phases between them are randomly distributed in the environment. The firing of each unit is conjunctively modulated by a spatial triangular grid and by its HD input.

A. Grid development with speed variation

Constant running speed is not a requirement of the model. To check this, we have simulated a rat with variable speeds, and left unchanged the method of choosing running directions. The trajectory of the simulated rat is then composed of epochs with positive or negative accelerations. The lengths of epochs are Poisson random numbers with mean length 3, roughly matching available behavioral data[6]. The speed at the end of each epoch is drawn from a truncated Gaussian distribution. A two-sided truncation is applied to keep the speed both positive and symmetrically distributed around the mean. The mean of the truncated Gaussian distribution is the same as the constant speed used in most simulations in this paper, i.e. 0.4m/s. The speed within each epoch is a linear interpolation between the speeds at the start and end of the epoch.

Regardless of different levels of speed standard deviation, grid fields form with high gridness scores (Fig. 5a), tight grid alignment and similar spacings (data not shown). Even for large speed standard deviation, 16.1 cm/s, i.e. 40% relative to the mean speed, the changes in mean gridness, mean standard deviation of grid orientations and mean gridness are all within $\pm$ one standard
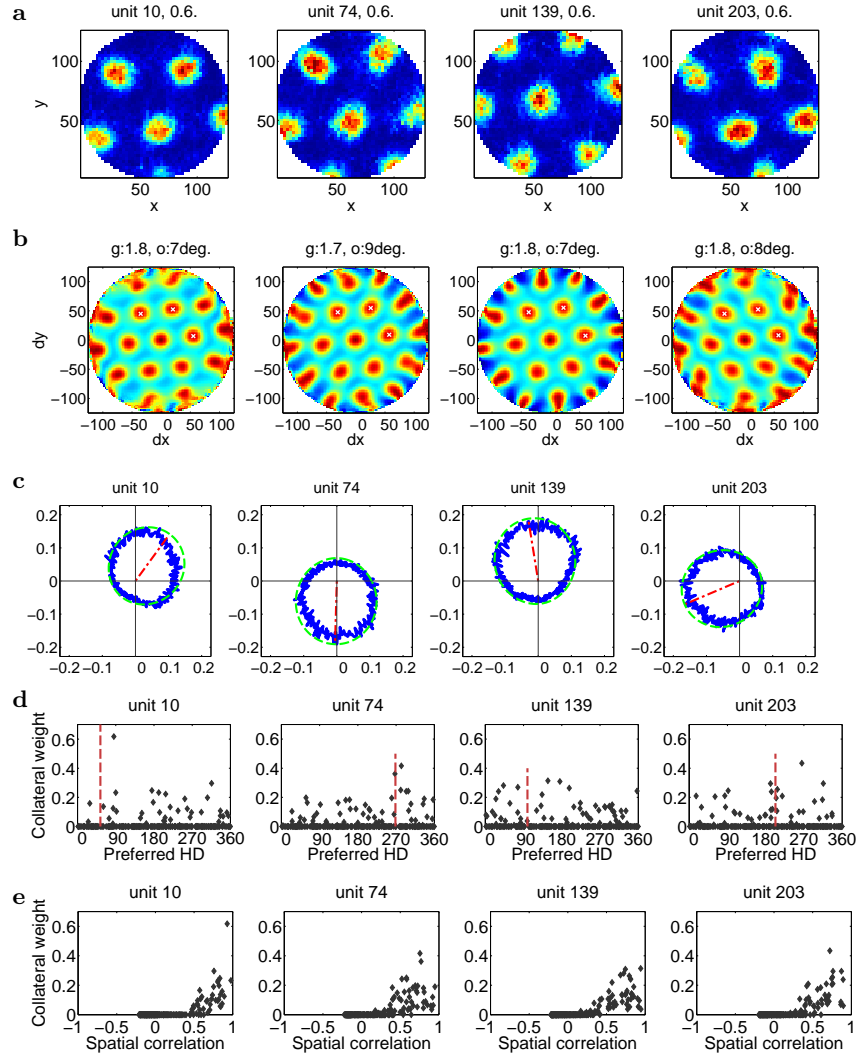
Figure 4: Collateral connections align grid fields. (a) Spatial firing rate maps of example units in a cylinder environment. Unit number and maximal firing rate (in arbitrary units) are indicated above each rate map; (b) Autocorrelograms of the maps shown in a. Gridness score and grid orientation (in degrees) are indicated above each autocorrelogram; (c) Angular firing maps (blue solid lines) of the same units as in a with, in green, the tuning curve of the HD input to each unit shown. The red dash-dot lines indicate preferred HD of the units; (d) Collateral weight W_{ik} as a function of θ_k. The broken line is θ_i; (e) Scatter plots between collateral weights and the spatial correlations of the fields between pre- and post-synaptic units.

B. Gradual grid development

Although grid formation is a gradual process, grids are expressed at relatively early stages of development[25,26]. In adult rats, grids appear after a first exposure to a novel environment, but need several days of experience to become stable[35]. Consistent with these findings, conjunctive units in our network also show early expression of grids, within a gradual formation process. As shown in Fig. 5b, grids can be observed already 20 minutes after the simulated rat has started to explore the envi-

ronment. With longer exploration, the grid fields become both more triangular and more coherently aligned to a common orientation, as quantified by the increasing mean gridness and the decreasing mean standard deviations of grid orientations (Fig. 5c,d). Grids stabilize after about 14 hours of continuous exploration. The mean spacing of the grids does not show big changes during development (Fig. 5e). Both the time scales of early grid appearance and of grid stabilization at the population level are comparable with experimental results, considering the time spent by a real rat in rest and sleep.
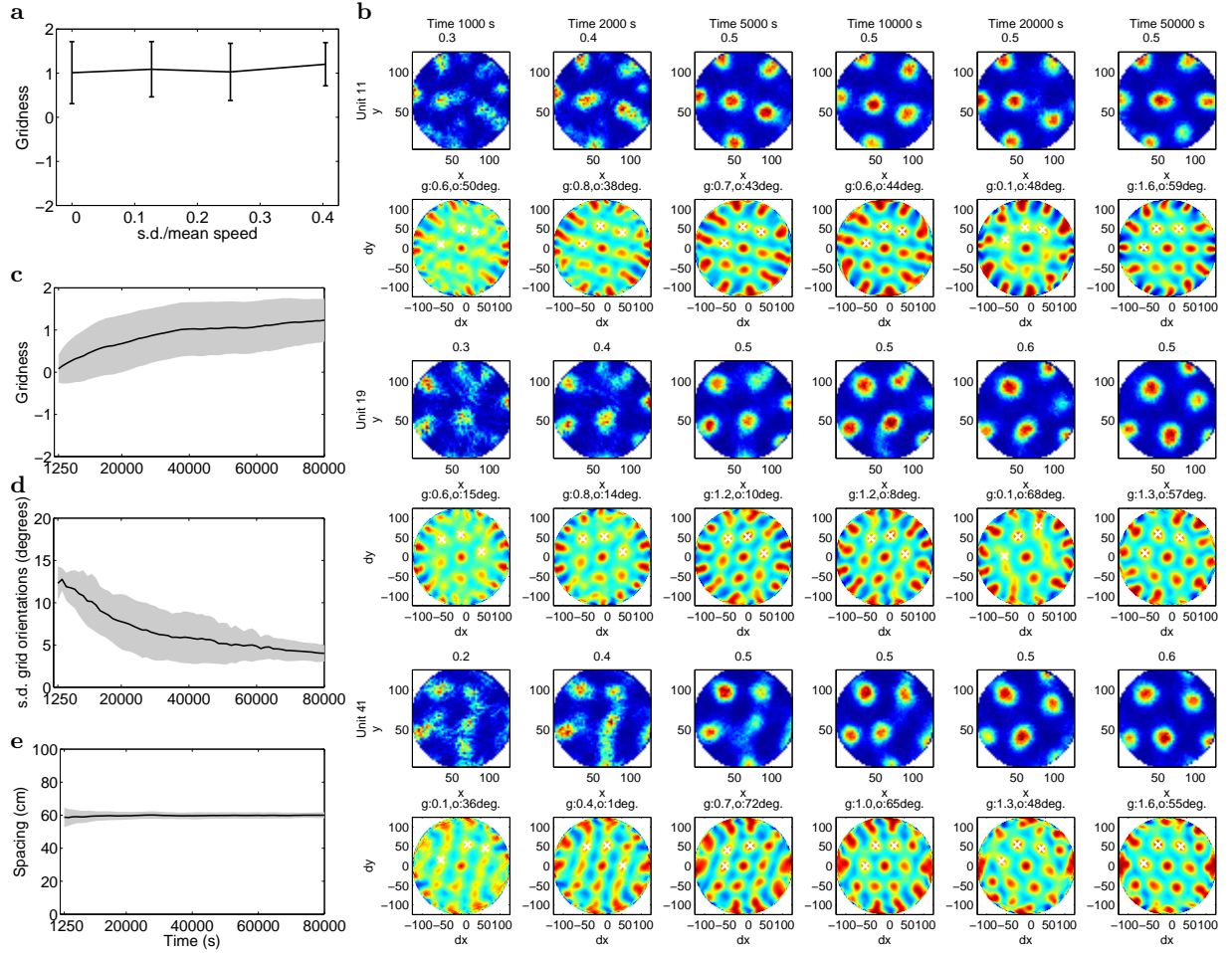
Figure 5: The variation in speed does not influence grid formation. (a) Grids show similar mean gridness scores at the end of simulations performed with different speed standard deviation relative to mean speed. Error bars indicate ± standard deviation across 5 simulations; (b) Grid fields appear in the early phases of the simulation and stabilize with more experience in the environment. Grid fields and the corresponding autocorrelograms are ordered in rows for three example units from the same network. Maps formed after the same amount of exploration are arranged in the same column; (c) Mean gridness increases with respect to the amount of exploration (speed standard deviation is 40% relative to the mean speed). Gray area is the ± standard deviation; (d) Standard deviation of grid orientations; (e) Mean spacing of the grids with gridness larger than 0.25.

IV. THE EFFECT OF THE SHAPE OF THE ENVIRONMENT

As shown in the previous section, grid fields align mutually to each other in cylinder environments. As cylinder has no preferred direction, and the common alignment emerging in one simulation bears no relation with the one emerging in another, leading to the expectation that different animals would show differently aligned grid units, if these were to form prevalently in cylinder environments. Rodent cages are usually rectangular, however, and a rectangle does have preferred directions. Is the orientation of grid units in different rats influenced by the shape of the training environment? The current section is devoted to this issue, namely, to the coherence of *grid orientation* across rats, i.e., in our case, across simulations. In contrast to the standard simulations in the previous section, we simulate virtual rats in a 125 cm ×

125 cm square environment, which leads to a (simulated) anisotropic exploration. We also vary the standard deviation σ_{RD} in running directions, which affects the degree of behavioral anisotropy.

Rats, and especially naïve ones, have a natural tendency to run along the walls of the environment. Therefore their trajectories would reflect the anisotropy of the enclosing perimeter. Our virtual rat in a square box has a probability higher than chance of following the walls, simply because of the random walk mechanism described earlier. While in the cylinder environment the trajectory is necessarily isotropic, following the walls in a square box will result in the emergence of four preferred running directions (Fig. 6 left column). The RD distributions for square environments exhibit four peaks centered around the directions parallel to the walls (Fig. 6 right column). In contrast, in a cylinder environment, the RD distribution is uniform (Fig. 3b). The non-uniformity of the RD
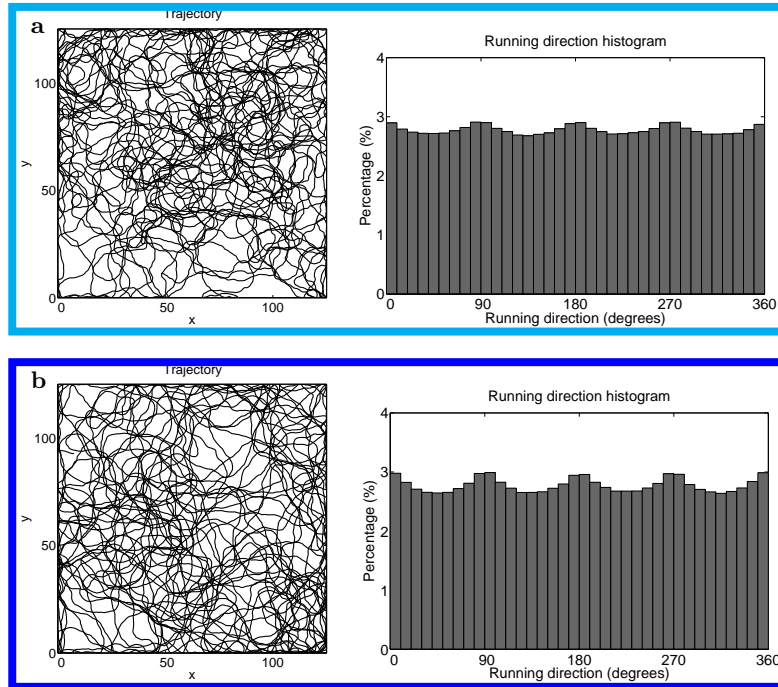
Figure 6: Square environments induce more running along the sides. The left panels show parts of typical trajectories. The corresponding distributions of RDs in simulations are depicted in the right panels. (a) The default standard deviation in RD, $\sigma_{RD} = 0.2$ radians; (b) A simulation with $\sigma_{RD} = 0.15$ radians, demonstrating stronger anisotropy in the trajectories.

distribution in square environments is stronger when the standard deviation in RDs is smaller (Fig. 6b).

In an environment with anisotropic boundary, if one running direction is preferred systematically, the network would associate the activity of conjunctive units more strongly to places following the preferred RD, and would effectively orient one of the grid axes along this direction, forcing the other two grid axes to follow. We would expect to see coherent grid orientation across rats in such conditions.

As in cylinder environments, gird fields show coherent alignment in individual simulations in square environments, irrespective of the RD standard deviation (Fig. 7a-c). To see any common orientation across simulations, we performed multiple simulations in square and cylinder environments with RD standard deviation $\sigma_{RD} = 0.2$ radians and in a square environment also with $\sigma_{RD} = 0.15$ radians. In each of the three conditions, 70 independent simulations were conducted with different seeds for the random number generator. Averaged and normalized over the 70 trials, the sum of the mean orientation distributions of the three grid axes in square environments shows a significant concentration at multiples of 30 degrees (Fig. 7d-e). That is, the common orientation of the grid fields is more likely to align along the walls of box environments. To quantify the coherence in orienting grids, a *cross-trial grid orientation coherence score* is defined for the one-dimensional mean orientation distribution of grid axes. The coherence score is computed in a similar way as the gridness score in Section III but assuming 30-degree periodicity, thus evaluating even and odd multiples of 15 degrees. Note that the range of the cross-trial grid orientation coherence score is [-2,2]. The cross-trial coherence of grid orientation in square environments is much larger than that in cylinder environments (Fig. 7d-e vs. f).

Results in Fig. 7 indicate that although the running-direction anisotropy associated with training in square environments does not influence the coherence in grid alignment, it may strongly modulate the coherence in grid orientation across rats. That is, running-direction anisotropy produces a single orientation for the aligned grids in each simulation, but two clusters of prevailing common orientations across simulations, at 30 (or rather, 90) degrees of each other, i.e. with one grid axis aligned to a wall of the square environment.

In the real system, two main orientations are observed in the same experiment (across grid units with different characteristic grid spacings[27]), not quite orthogonal to each other, but also not aligned to the walls of the testing environment. It could well be that these prevailing orientations reflect other environments, where rats were caged during development. A rat who develops in a square or rectangular cage, however, is expected to experience not only RD anisotropy, as tested in the model, but also speed anisotropy. To disentangle a potential contribution of the latter, in the following section we introduce speed anisotropy, and compare its effects with those of running-direction anisotropy.
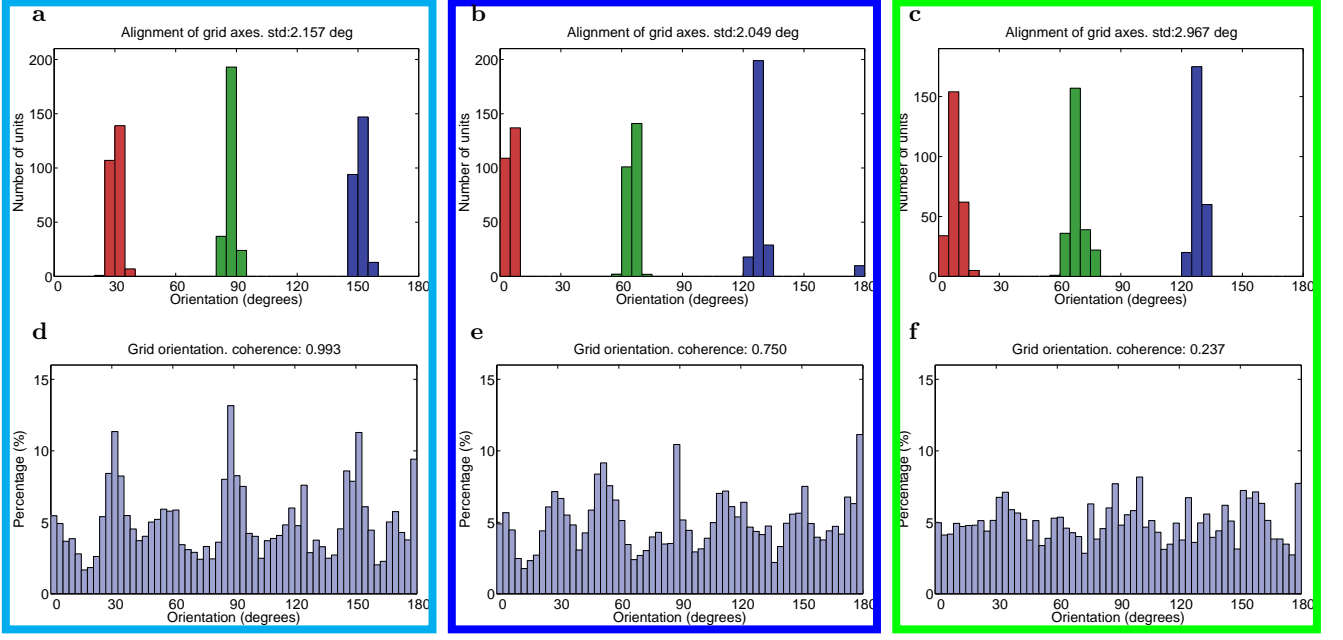
Figure 7: Square environments tend to orient the aligned grid fields along the directions of the sides, compared with cylinder environments. (a-c) Grid fields align to each other in square environments (left column: $\sigma_{RD} = 0.2$ radians, middle column: $\sigma_{RD} = 0.15$ radians) and in a cylinder environment (right column: $\sigma_{RD} = 0.2$ radians) in three individual simulations. The alignment coherence score, i.e. the mean standard deviation (in degrees) of the orientations averaged over the three grid axes, is indicated at the top of each panel. The same data in Fig. 3e is shown in c again; (d-f) The sum of the orientation distributions of the three grid axes averaged over 70 simulations. In square environments, the orientation distribution of the three grid axes shows periodic clusters, which are distinct from the random fluctuations of the average orientation distribution. The number at the top of each panel indicates the coherence of grid orientation.

V. SPEED ANISOTROPY

Having observed that grid orientation is influenced by the non-uniformity of running directions, in this section we focus on another factor in the trajectories of the simulated rat, namely running speed, and investigate whether the shape of the grids can be influenced by an anisotropic speed distribution in exploration behavior, even with no anisotropy in the boundary conditions (i.e. in cylinder environments). In this section, the simulations are identical to those in Section III, except that the virtual rat explores the cylinder environment running faster in four preferred directions, as visualized in Fig. 8

$$v^t = v_s[q + (1-q)\frac{|\sin(\omega^t)|^3 + |\cos(\omega^t)|^3 - 1/\sqrt{2}}{1 - 1/\sqrt{2}}]. \quad (11)$$

Here ω^t is the current running direction of the rat. The speed now can go from a minimum of $v_s = 24cm/s$ to a maximum of $v_s = 40cm/s$ depending on the running direction. $q = 0.6$ is the ratio between the two speed extremes.

As in the standard simulations in Section III, the RD distribution of the trajectory is quite uniform (Fig.9a,b). In the RD distribution there is a tiny over representation of running directions parallel to the diagonals. This tiny distortion is originated by the fact that when the rat is
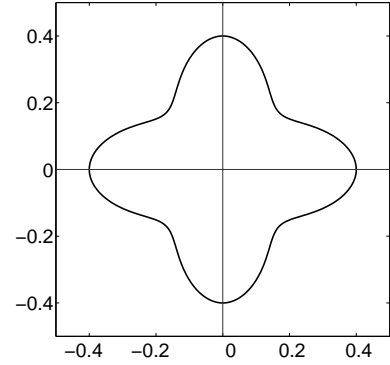

Figure 8: Polar plot of the speed of the simulated rat with respect to running directions. The maximal speed is the default speed $0.4m/s$, and the minimal speed is 60% of v_s.

running faster it takes a shorter time for it to reach the boundary, where it is forced to turn. However, the effect appears much smaller than that due to the square shape of the environment in Section IV, and its direct influence on grid alignment is negligible.

Fig. 9j shows the spatial maps as well as the autocorrelograms of four example conjunctive units from the network. Similar to the case with constant speed, grids align with each other, and the phases are kept broadly distributed in the environment (Fig. 9d-f). However, the

average gridness score is somewhat lower as compared to the case with constant speed (Fig. 9c). The reason is that, with speed anisotropy, the grids are distorted, showing grid axes with non-equal lengths, as can been seen carefully in Fig. 9j. Among the three axes of a grid, the axis that passes through the maximum farthest from the origin is identified as the long grid axis. In our simulations with speed anisotropy, conjunctive units tend to develop maps that share the same grid axis as the long one (in the simulation of Fig. 9e this corresponds to the axis with orientation around 100 degrees), indicating that grid distortion happens predominantly along one direction. Another way to quantify this distortion, introduced in[27], is to fit an ellipse that passes the six maxima close to the center of the grid autocorrelogram (white curves in Fig. 9j). These ellipses deviate from perfect circles, indicating modified grid structures. We find that such ellipses are roughly aligned in each map with the corresponding long grid axis, laying within 30 degrees of it (Fig. 9g-h). The flattening of an ellipse is measured by ellipticity, i.e. the ratio between the major and minor axes. Note that the ellipticity of a perfect circle is 1. In the simulation, the median of the ellipticity distribution of the grids is about 1.15 (Fig. 9i).

Averaged over 70 independent simulations in cylinder environments, the orientation of the long grid axis concentrates along the preferred directions of the speed profile, i.e. 0 and 90 degrees (gray bars in Fig. 10a). The other two grid axes only broadly orient to 60/120 or 30/150 degrees respectively (light blue bars in Fig. 10a), with low coherence of grid orientation. In contrast, with constant speed in cylinder environments, both the average orientation distribution of grid axes and of the long grid axis are fairly uniform, reflected in even lower coherence score of grid orientation (Fig. 10d).

With speed anisotropy in cylinder environments, we observe that grid maps are more elliptical than those developed from exploration with constant speed (Fig. 10c vs. Fig. 10f). The difference is rather subtle, however, as the fluctuations in the length of the grid axes, and our choosing always the major axis of the best fitting ellipse, whichever it is, obviously produce an ellipticity measure distributed above 1, even with constant speed. The effect of anisotropy is more salient in orienting the ellipses, Fig. 10b, around the preferred directions of the speed profile, a clear difference from the uniformly distributed ellipse orientation of simulations with isotropic speed, Fig. 10e. The difference in orienting ellipses can be quantitatively measured by defining *cross-trial ellipse orientation coherence score* for the one-dimensional average distribution of ellipse orientations. The cross-trial ellipse orientation coherence score is calculated the same way as the cross-trial grid orientation coherence score, except that 90-degree periodicity is assumed, instead of 30-degree periodicity. With speed anisotropy in cylinder environments, the cross-trial coherence in ellipse orientation is as high as 1.533. However, with constant speed, still in cylinder environments, the corresponding coher-

ence score is close to zero, indicating a uniform ellipse orientation distribution. Speed anisotropy orients ellipses rather loosely, with a width at half-peak around 30 degrees. It is reflected also in which of the three grid axes tends to be the longest, but without apparently forcing a rigid orientation of the grids themselves, Fig. 10a.

Fig. 11 shows a summary of the differences between speed anisotropy and RD anisotropy, once coupled with RC interactions, with respect to three measures, namely, cross-trial coherence in grid orientation, cross-trial coherence in ellipse orientation, and mean ellipticity. One can see that both anisotropies slightly distort the shape of grid maps into a somewhat more pronounced elliptical arrangement of the firing fields. Speed anisotropy induces on average more distortion (with our parameters), but the average effect is overshadowed by the large variability from unit to unit, in what is an asymmetric ellipticity distribution with a long tail. The qualitative difference is in what common orientation emerges: speed anisotropy loosely orients *ellipses* toward the fast directions (larger cross-trial coherence in ellipse orientation), but it does not enforce a tight grid orientation, across simulations, on the aligned grids (dark circle in Fig. 11a-b). In contrast, RD anisotropy orients *grids* toward preferred RDs (larger cross-trial coherence in grid orientation, white and gray squares in Fig. 11a-b), but not ellipses. When combined with RD anisotropy, speed anisotropy prevails (the dark square in Fig. 11a-b). This is likely because speed aniso-tropy induces the long grid axes to lie along either 0 or 90 degrees, and it constraints the lengths (shorter) and orientation (not at 60 degrees) of the other two grid axes; with RD anisotropy, the arrangement of the fields maybe more dependent on the details of the trajectory taken in each simulation, which does not appear to restrict much the orientation of the ellipses, seemingly leaving the grids free to orient themselves. We may conclude that the shape of the environment and speed anisotropy cause apparently similar, but subtly distinct effects.

A. Grid alignment without collaterals

In the simulations above, collateral interactions among would-be conjunctive units tend to align their developing grid fields along common axes. We have considered two additional factors that influence the alignment, the shape of the environment, if different from cylindrical, and anisotropy in running speed. The latter factor, in particular, enhances the ellipticity of the resulting grids, and affects their ellipse orientation. Its effects are however secondary, in the simulations, to the primary effect produced by the collaterals. It is important to note that ellipticity and some degree of common orientation, however, can be produced also by speed anisotropy on its own, in the *absence* of collaterals. This can be understood by considering a simple abstract model, which extends the one earlier considered in[23] to account for the development of grid fields at the single unit level. The model
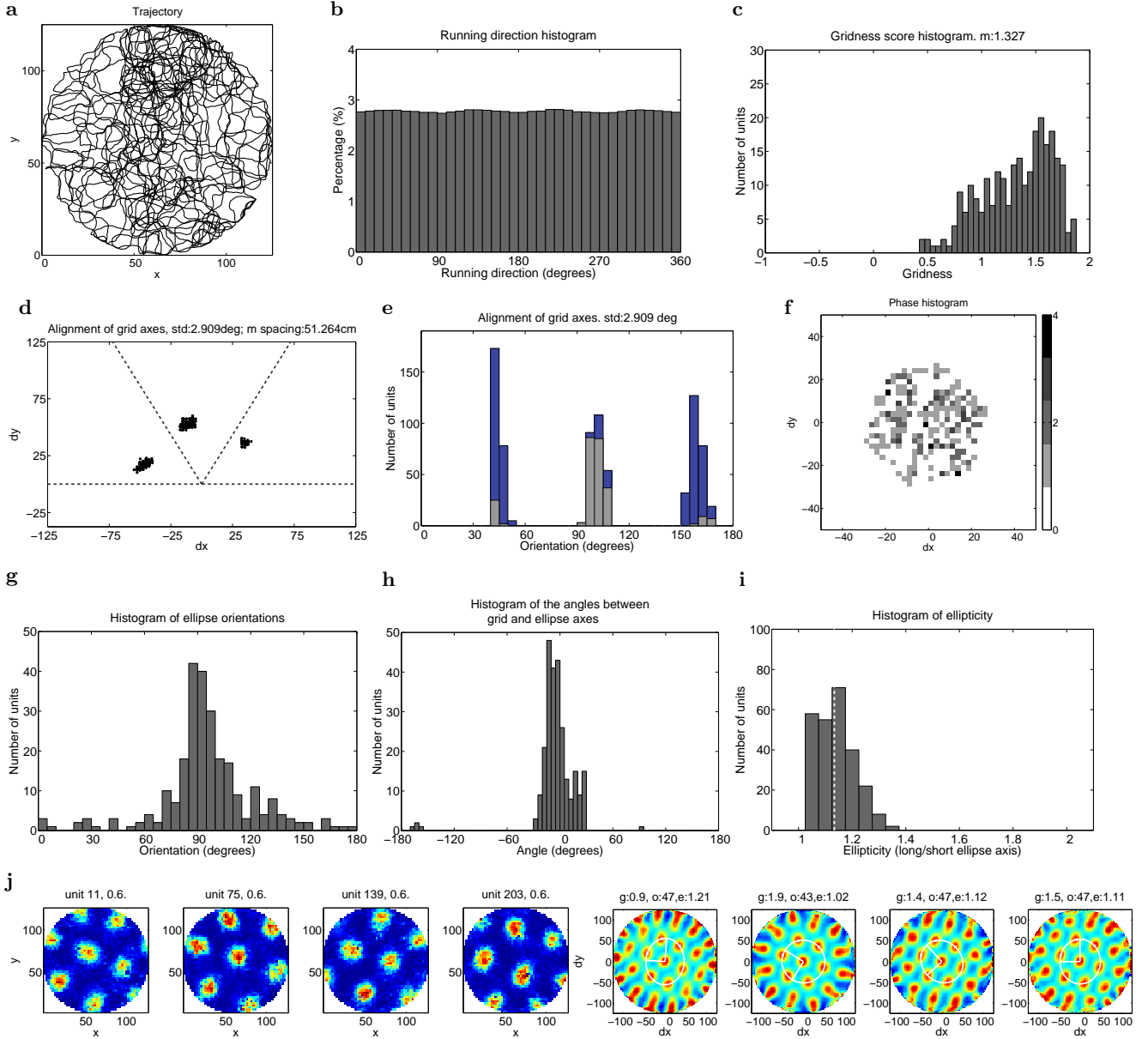
Figure 9: Grid alignment during exploration in a cylinder environment with anisotropic speed. (a) A trajectory of the rat with 2×10^4 steps in a cylinder environment; (b) RD distribution of an entire trajectory in a 8×10^6-step simulation; (c) Histogram of the gridness scores of all conjunctive units in the simulation; (d) The scatter plot of the locations of the three peaks found in autocorrelograms; (e) Histogram of the orientations of the three grid axes. Shown in front in gray is the orientation histogram of the long grid axes. Indicated at the top of the panel is the coherence score in grid alignment (mean standard deviation, in degrees, of the orientation averaged over the three grid axes), similar as the coherence in grid alignment in the cylinder environment with constant speed; (f) The histogram of spatial phases (again, relative to that of the best grid); (g) Histogram of the orientations of the major axes of the ellipses determined from each autocorrelogram; (h) Histogram of the angles between the long grid axis and the ellipse major axis; (i) Histogram of the ratios between the major and minor axes of ellipses. The white broken line indicates the median; (j) Examples of the fields of conjunctive units (left) and the corresponding autocorrelograms (right). The white curves show the ellipses determined from the three maxima found in autocorrelograms. Unit number, and maximal firing rate (in arbitrary units) are indicated above each rate map. Gridness score,orientation (in degrees) and ellipticity are indicated above each autocorrelogram.

describes a single unit in a very large environment, which then for all practical purposes has no shape. Hence the abstract model focuses on the effect of speed anisotropy, which is modeled similarly as in the simulations, extri-

cating it from the other two factors, the presence of collaterals, and the shape of the environment. The model, described in the Appendix, leads to a relation between the degree of anisotropy and the degree of ellipticity, and
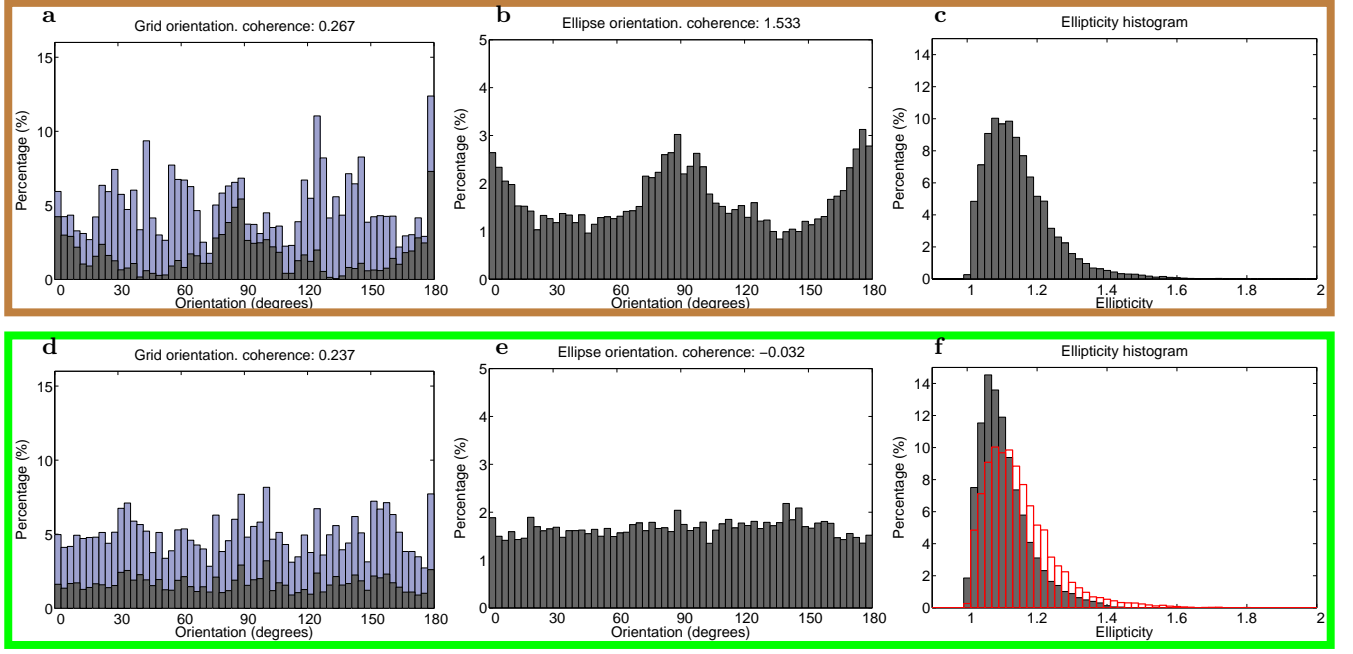
Figure 10: Speed anisotropy in cylinder environments distorts the shape of the grids (top row) as compared to exploration with isotropic speed (bottom). (a,d) Average orientation distribution of long grid axes (gray) plotted in front of the average orientation distribution of the three grid axes (blue). The coherence of grid orientation is indicated at the top of each panel. The blue bars in d are the same measure as the distribution shown in Fig. 7f; (b,e) Average orientation distribution of the major axes of the ellipses. At the top of each panel, the coherence in ellipse orientation is noted. It is calculated similarly as the coherence in grid orientation, but assuming 90-degree periodicity instead of 30-degree; (c,f) Average distribution of ellipticity. The distribution in c is plotted again as the red empty bars in f.
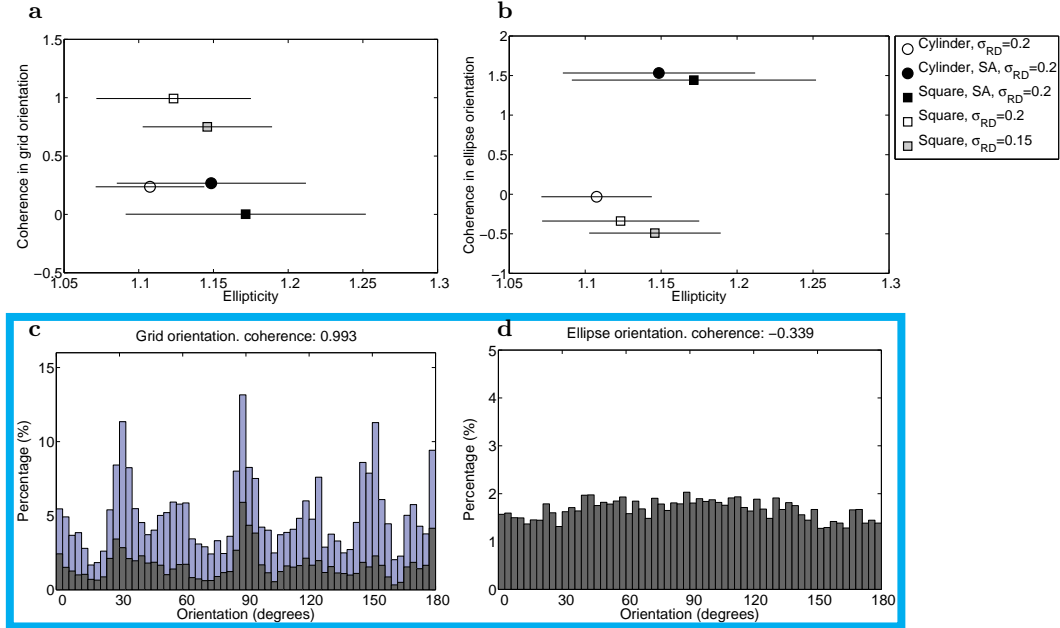

Figure 11: Speed anisotropy and RD anisotropy have different effects on grid shape and orientation. (a,b) Five situations are compared with respect to coherence in grid/ellipse orientation and to ellipticity, based on the average over 70 independent simulations. The horizontal lines indicate ± one standard deviation in ellipticity. The two panels share the same legend to the right; (c) The average orientation distribution of the long grid axes (gray, in front) and the average orientation distribution of all three grid axes (blue, in the back) for the simulations in square environments without speed anisotropy, the same data as shown also in Fig. 7d. The coherence in grid orientation is noted at the top of the panel. (d) The average distribution of ellipse orientation, also for the simulations in square environments without speed anisotropy, with the corresponding coherence at the top.

it indicates two orientations, at roughly 90 degree of each other, of the grids and of the ellipses that best match the assumed quadrupole anisotropy. Hence it predicts that grids and ellipses would have one of two common orientations even in the absence of collateral interactions.

The effect predicted by the analytical model is reproduced in *ad hoc* simulations. Fig.12 shows that in the presence of speed anisotropy only, without collateral interactions, grid units develop with an enhanced ellipticity, and with one of the grid axes preferentially aligned along one of the two orthogonal directions with higher mean speed (in the simulation of Fig. 12, this happens to be the one at 90 degrees). The variance in ellipse orientation is considerable, larger than what is observed with speed anisotropy *and* collateral interactions (Fig. 12b vs. Fig. 9g). However, the degree of ellipticity is similar, irrespective of the existence of collateral interactions (Fig. 12c and Fig. 9i). The first element that is missing, without collateral interactions, is crucially the tight alignment of the grids with each other. This is because, in our model, collateral interactions align the grids with each other through mutual iterative convergence, whereas without collaterals the alignment only reflects single unit adaptation to what is effectively a broad shallow valley in a free energy landscape. We present the abstract model in the Appendix for clarity, but we believe it to be unlikely that, in the absence of collateral interactions, speed anisotropy alone can establish a common grid orientation with the tight alignment seen in experimental data. The second element that is missing without collateral interactions is the invariance in the relative phase of grid maps across multiple environments, which we will examine in the next section.

VI. GRID REALIGNMENT IN MULTIPLE ENVIRONMENTS

Until now, we have only considered grid alignment in a single environment. Under dramatic environmental changes, both the hippocampal and entorhinal neurons develop new maps, in a process called global remapping[28]. While hippocampal place fields seem to shuffle randomly during global remapping, grid cells behave in a population-coherent manner. The new maps preserve the same relative phases as the old maps, and maintain a common grid orientation, not necessarily the same one as before[29]. In this section, we address the question of how grid fields may possibly align in multiple environments.

We simulate the network in two different cylinder environments, with identical conditions as in Section III, except for the number of units. The total number of conjunctive units is 500. The total number of place units is 900. In each environment, 500 of the place units are active. The number of place units that are active in both environments is 100, i.e. 20% of the active place units in one environment. The place fields of the place units in the first environment are completely different from those in the second environment. However the preferred HDs of the conjunctive units are the same across environments. The training is interleaved in the two environments, with 2000 epochs in each environment, 3000 steps in each epoch. This leads to 1.2×10^7 training steps in total.

Fig. 13a,f compares the spatial fields and corresponding autocorrelograms of four example conjunctive units in two environments. In both environments, the grids are aligned, but to different orientations (Fig. 13b,g). The angular offset of the grids in the second environment relative to the grids in the first environment is about 9 degrees counterclockwise. The spatial phases of the units relative to the best grid in each environment do not show any clear pattern or clustering (Fig. 13c,h), as in the simulations with a single learned environment.

The distribution of ellipse orientations is quite different across environments (Fig. 13d,i), but the ellipticity distribution is comparable, with a similar median at 1.16 (Fig. 13e,j). This value is similar to that of the simulations with speed anisotropy. This is because 20% of the place units are active in both environments but in randomly different positions. Since conjunctive cells need to maintain population coherence, this randomness acts as a source of noise, imposing additional constrains on conjunctive units that move their maps away from perfect gridness. This can be seen as a third source of ellipticity. The other two sources discussed in previous sections, i.e. speed anisotropy and RD anisotropy, introduce ellipticity during grid development by breaking the symmetry of the trajectories of the simulated rat. The ellipticity caused by grid realignment in multiple environments is due to the overlap of the population codes of contextual information, and is imposed by the structure of the network.

We then determine whether the relative spatial phases of the units are kept invariant across environments. For this we select a sub-population, by taking into account only units that have good grid maps in both environments (gridness score > 0.75). The firing map in environment B of each unit in the population is rotated counterclockwise at multiples of 4.5 degrees and cross-correlated with the corresponding map in environment A, as in[29]. The cross-environment crosscorrelograms thus obtained are averaged over the population, to get the mean cross-environment crosscorrelograms for this sub-population.

The mean crosscorrelogram of two groups of grid maps shows whether the grid maps are related by the same shift transformation. When two grids have the same orientation and spacing, the crosscorrelogram between them shows grid structure. In addition, the relative spatial shift between them determines the direction and distance that the corresponding crosscorrelogram is displaced away from the center. Cross-correlating one group of grids, which have the same orientation and spacing but distributed spatial phases, with a second group of grids, which are a shifted version of the first group for the same amount, will result in a set of crosscorrelograms with grid
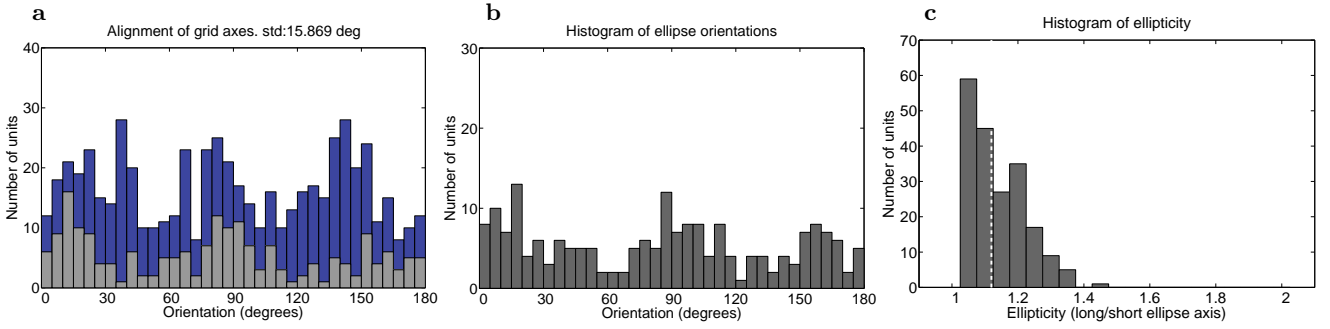
Figure 12: Grids show a weak tendency to align due to speed anisotropy alone, without collateral connections in the network ($\rho = 0$). (a) Histogram of the orientations of the grid axes (blue, in the back) together with the histogram of the long grid axes (light gray, in the front). The coherence in grid alignment is low (the mean standard deviation, in degrees, of the orientation averaged over the three grid axes is shown at the top of the panel, and it is much higher than with collaterals, shown in Fig. 9e); (b) Orientation histogram of the major axes of the ellipses; (c) Histogram of ellipticity.

pattern and identical offset away from the center. The offset is just the common shift between the two groups. Adding the crosscorrelograms between the grids in the two groups preserves a clear grid structure. If a common rotation of multiples of 60 degrees is applied, instead of a common shift, the second group of grids still have the same orientation as the original grids, and each individual crosscorrelogram still keeps a grid structure, but the phases are changed differently for each grid (Fig.15), producing crosscorrelograms with peaks appearing in distributed locations. Thus, the mean crosscorrelogram of the two groups does not maintain a grid pattern. For a common rotation other than multiples of 60 degrees, the rotated grids do not align to the same orientation as the original grids any more, making it impossible for the mean crosscorrelogram to be a grid. Therefore, for two groups of grids, the mean crosscorrelogram appears to be a grid only when there is a coherent shift between the two groups.

Fig.14a shows that only with a counterclockwise rotation close to 9 degrees does the mean cross-environment crosscorrelogram show a grid structure. This is the angle that aligns the grid axes of both environments in Fig.13b,g. Fig. 14a thus indicates that after the angular offset is counter-balanced, the spatial shifts between the grid fields in the two environments are the same (Fig. 14b). For other rotations, in contrast, the spatial shifts are different across units, giving rise to an increasingly flat mean cross-environment crosscorrelogram. The absolute angular offset between the grid orientations in two environments is not important. Note that our simulations have fixed HD selectivity across environments, whereas in rats the preferred HDs of both conjunctive cells and HD cells appear to rotate coherently by the same amount in different environments[6]. Therefore the angular offset between the two sets of grid maps in our simulations can be considered as the angular shift after correcting for the HD selectivity shift across environments.

Fig.14c shows the "gridness" scores of the mean cross-environment crosscorrelogram with respect to the amount of rotation when a different number of units from the network is included in the analysis. The "gridness" score defined here is similar to the one in the method described in Section III. The difference is that the ring circling the six maxima is centered on the closest maxima from the center of the crosscorrelogram, and that the difference in covariance instead of the correlation coefficient is used to define gridness. This is because most of the mean cross-environment crosscorrelograms are quite flat, giving unstable correlation coefficients when normalizing with small variance. As can be seen in Fig.14c, with 20 units as a population, it is already possible to infer the angular mismatch between the fields in two environments. When only one unit is considered, the gridness of the cross-environment crosscorrelogram is roughly periodic with period 60 degrees. This indicates that with the population code expressed by grid cells it is possible to keep a unique orientation correspondence across environments in spite of the orientation symmetry of the grids, and to differentiate locations in an environment by the relative spatial phases of the grid fields. To represent places on a larger scale, multiple grid spacings are needed[37–40].

VII. DISCUSSION

Several conclusions can be extracted from these simulations. In the first place, we have described how alignment can be achieved in the model through recurrent collaterals, as long as they are associated with HD information. Secondly, we have shown that running-direction anisotropy in behavior, induced by the shape of the environment, is enough to distort the grid maps into ellipses, but it has the side-effect, of orientating one of the grid axes with one of the walls. At the same time, whether because too weak or localized, this kind of anisotropy fails to align the distortion, measured by the long grid axis or ellipse direction, into a common orientation. We
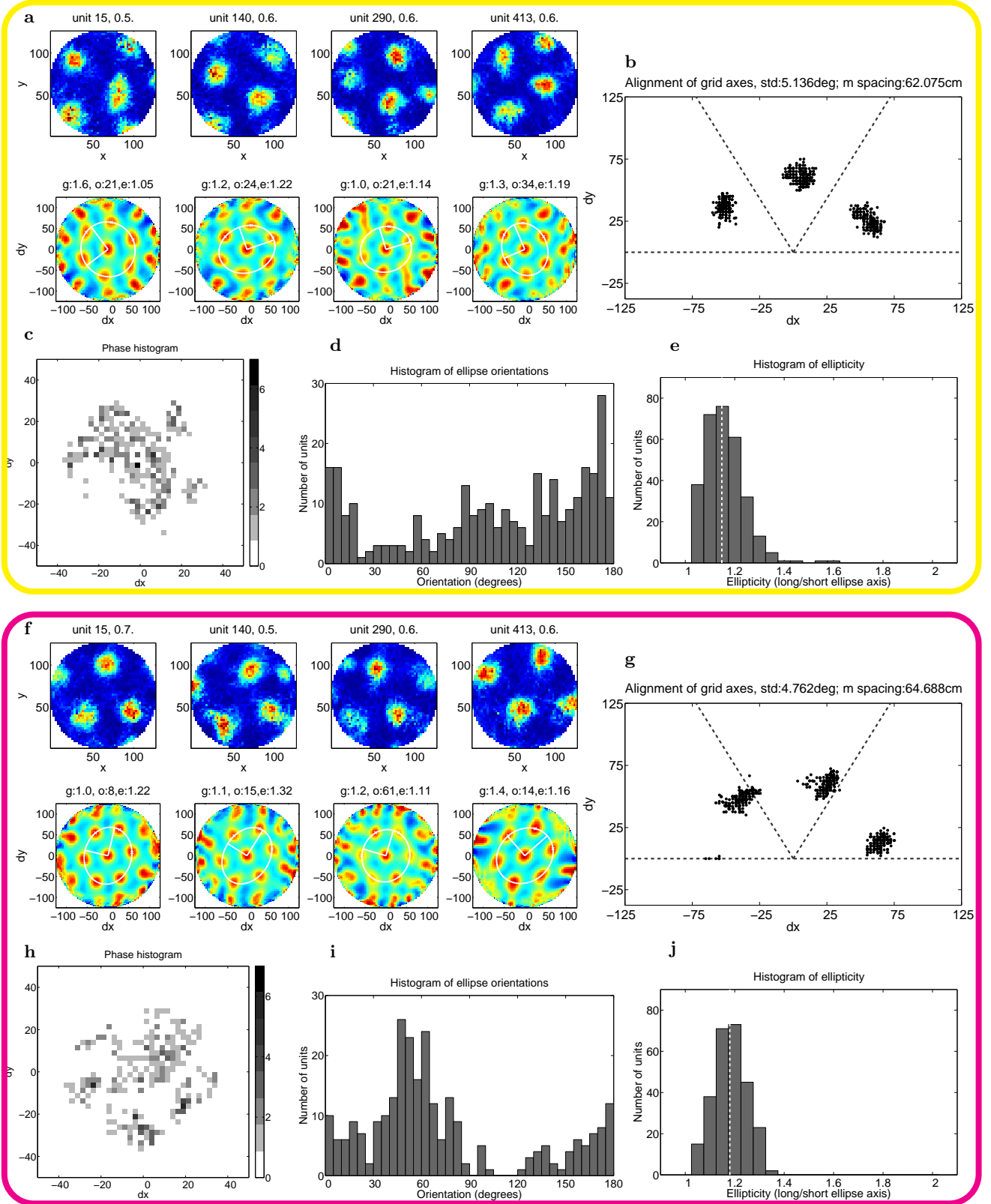
Figure 13: Grid realignment in environment A (top) and environment B (bottom). (a,f) Examples of the spatial fields and autocorrelograms of four different units in the two environments; The unit number and maximal firing rate (in arbitrary units) are indicated above each rate map. The gridness score, grid orientation (in degrees, between the x-axis and the nearest grid axis) and ellipticity are indicated above each autocorrelogram; (b,g) Scatter plots of the three peaks found in autocorrelograms. The angular offset of the grids in environment B relative to the grids in environment A is 9 degrees clockwise. Noted at the top of the panel are the mean spacing and coherence score in grid alignment (mean standard deviation, in degrees, of the orientations averaged over the three grid axes); (c,h) Two dimensional histograms of spatial phases relative to the best grid in each environment. Each spatial bin represents a $2.5cm \times 2.5cm$ area; white indicates zero units with a phase in that spatial bin, with more units indicated by progressively darker gray; (d,i) Histograms of the ellipse orientations; (e,j) Histograms of the ellipticity, i.e. the ratio between the length of the major and minor axes of an ellipse, across all units.
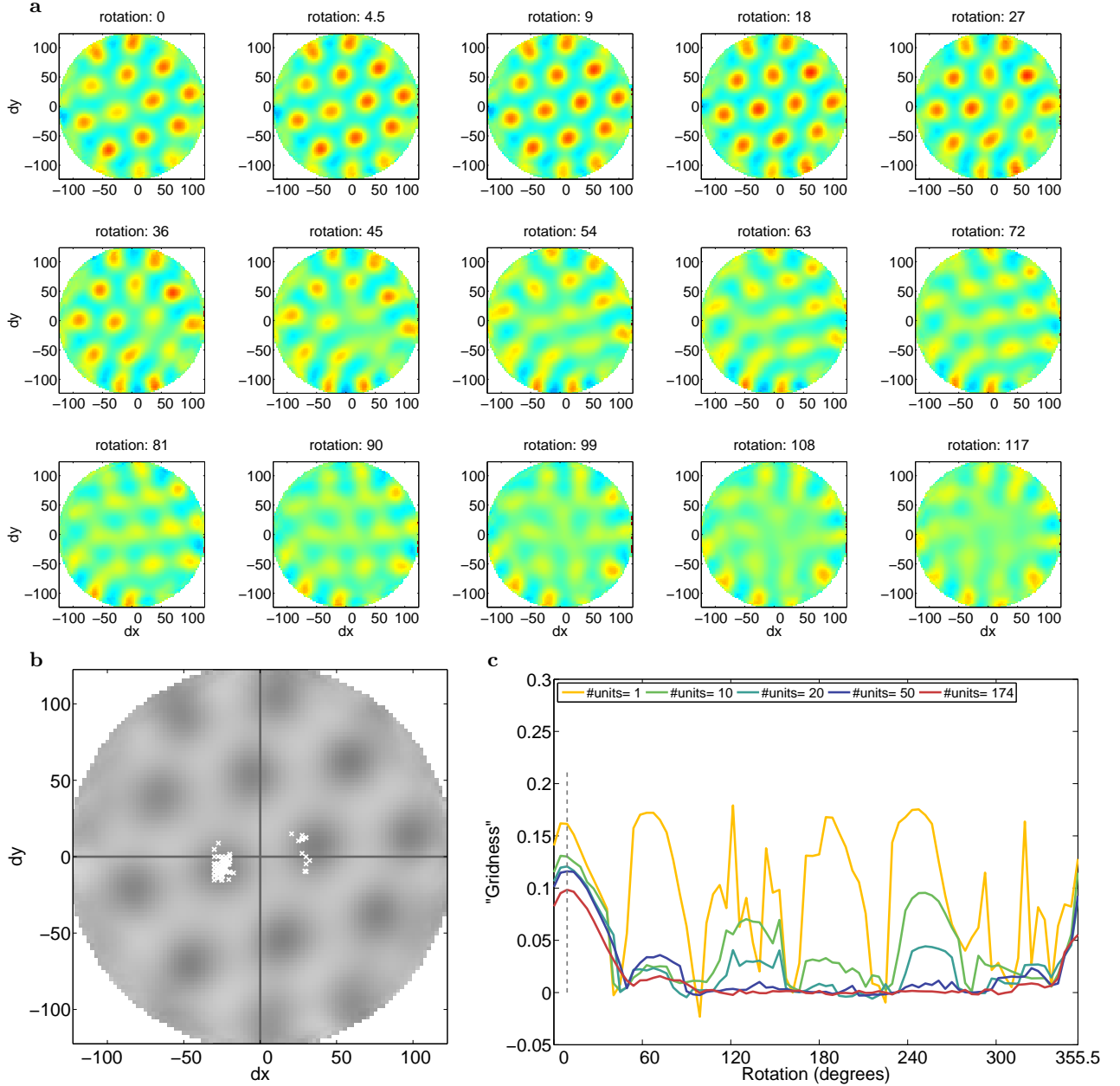
Figure 14: The phases of the grids in the two environments rotate and shift together. (a) The mean cross-environment crosscorrelogram of the population after the optimal counterclockwise rotation of 9 degrees is shown, together with some other rotations. The rotation angle is noted on the top of each panel. The mean cross-environment crosscorrelogram shows the best grid structure in the case of the specific rotations that (roughly) match the angular offset between the grid axes in the two environments. The Mann-Whitney test[36] shows that the mean cross-environment crosscorrelogram after 9-degree rotation has larger mean than the crosscorrelogram after zero-degree rotation (p=0.0015); (b) After the optimal rotation, the phase shifts of the same units concentrate in a small region with size similar to the field of a grid. Note that the phase shifts of a small number of units fall in a different area, due to the periodicity of the grid and to small fluctuations in the peaks of the individual crosscorrelograms. The gray crosscorrelogram shown in the background is the mean cross-environment crosscorrelogram after the optimal rotation, as shown in a; (c) The "gridness" of the mean cross-environment crosscorrelogram when the population set is increased from a single grid (# units 1) to the whole selected population. The gray broken line indicates the optimal rotation leading to the highest gridness of the cross-environment crosscorrelogram.
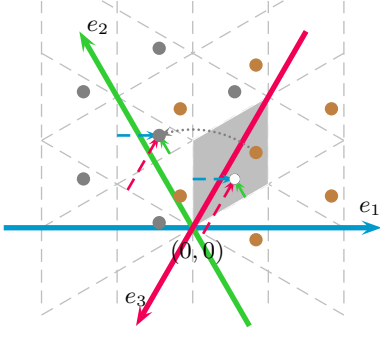
Figure 15: Rotating a grid map by 60 degrees results in a non-linear spatial phase transform. The spatial periodicity of a grid can be decomposed into the periodicity along the projection axes e_i, defined perpendicularly to the three grid axes. The orientation of and the period/wavelength along the third projection axis is constrained by the other two projection axes. The spatial phases of grids with the same spacing and orientation can be uniquely represented by points in a rhombus (area in gray). The sides of the rhombus are parallel to two of the grid axes, and the length of the side is equal to grid spacing. The spatial phases are only distinguishable in a rhombus area, because the rhombus is exactly one period length if projected onto two of the projection axes, e_1 and e_2. Points outside the rhombus are equivalent to those in the rhombus, modulo the wavelength along each projection axis. A grid with non-zero phase (shown in brown) is rotated by 60 degrees (indicated by the dotted arc) around the origin. The phase of the rotated grid (in gray) is equivalent to the white dot in the gray rhombus, since the coordinates (dashed arrows) of the two points along the projection axes are the same, modulo the wavelength. The phase change caused by the rotation depends on the initial phase of the grid.

have then explored the effect of running speed anisotropy, which has a milder influence on grid orientation but a stronger one on the orientation of the ellipses. In this scenario, recurrent collaterals have the effect of amplifying through population coherence what could be seen as a distortion of the idealized grid map already at a single cell level. This, however, may be regarded as a side effect of the function of recurrent collaterals in aligning the grid maps. Such a function may be useful, we hypothesize, in order to achieve global remapping with an invariant structure of relative phases, which, as we show in the last part of the paper, our model reproduces when the virtual rat explores two different rooms.

Whether grid alignment results from network interactions or as a consequence of a single unit process is a very interesting and complex question. In this respect, the literature presents opposing views.

In models of interference between multiple oscillators, the grid maps are the result of information processing within a single neuron[16–19]. Each grid cell is then independent, and the common alignment comes from the fact that they all share the same path integration inputs. The different spatial phases are achieved by anchoring each grid cell to a different set of spatial inputs. This selective anchoring is in a way analogous to what happens with

our place inputs, but it plays in the oscillator-interference models the minor role of correcting the errors of path integration, which is the main source of information used to construct the grid maps.

A completely opposite view to the single unit perspective comes from the attractor models[8–12]. In these, grid maps do not emerge individually, but as an emergent property of the population, so that grid maps and grid alignment are intrinsically inseparable. In our model, as shown by Kropff and Treves (2008), grid cells with no recurrent collaterals can develop grid maps with independent random orientations if they are left to interact weakly through inhibition, a situation that reminds us of the anatomy of layer II of medial entorhinal cortex. The introduction of collateral connections with the necessary head direction information to disambiguate firing sequences, as suggested earlier and proved here, can achieve grid alignment, as observed in experiments, thus offering a possible explanation of why these elements are found together in layers III to VI and are both absent in layer II. In our model, the common grid orientation lies on the attractor manifold created by the collateral interactions. This attracting property is a feature shared by the attractor models. In the latter, however, the collateral interactions have the full job of creating both grid maps and grid alignment.

The alignment of grid cells in layer II is perhaps inherited from the deeper layers, which send strong projections to superficial layers[41]. This hypothesis, not developed in the present work, represents an interesting direction of modeling research, which could focus on how conjunctive cells and grid cells may develop together in different layers, and perhaps even self-organize into laminar structures[23,42].

From the point of view of our model, the best way to address the question of a single unit versus a population origin of grid alignment is to interfere with the head direction source of information, assuming with some plausibility that this signal is not directly generated in the entorhinal cortex. Without head direction information, sequences in the grid cell network would be ambiguous. It would be interesting to study experimentally the effects that this kind of manipulation would produce in layers III to VI, where we assume that collateral connections work hand by hand with head direction information.

The functional difference between feed-forward connections and collateral connections can be appreciated in global remapping. In each environment, only a subset of place units is active. Feed-forward connections therefore relay environment-specific inputs to the conjunctive units. The collateral connections, however, signal the interactions between conjunctive units by the environment-independent geometrical structures they encode. With the same feed-forward and collateral connections, the network is able to provide an efficient metric encoding of space for all environments, since the emerging grid maps are essentially the same set of grids, almost identical after a unique rotation and translation. A natural ques-

tion to ask is how the collateral weights can be learned. Addressing this important issue will be very instructive for the attractor models as well, and therefore requires a separate article to be elaborated in depth.

As we show in the present work, a weaker form of alignment of grid cells can be produced with independent single units, if the behavior of the rat includes some form of anisotropy. This alignment, however, would imply no population coherence in relative phases with global remapping. Still, anisotropy might be the reason why grid maps present distortion or ellipticity. This is not the first time that behavior is proposed to influence the firing of grid cells. In the hairpin maze experiment[43], rats covered a two dimensional space but walking through a series of corridors in such a way that the behavioral constraints were those typical of exploration in one dimension. As a result, the grid cells showed a fragmented map with several one-dimensional patches instead of the canonical triangular structure. In our model, behavior alone creates distortion in the symmetry of grid maps, which is amplified and made coherent by recurrent collaterals. In a way, the distortion observed by Stensland and colleagues (2010), coherent at the population level, might be the price that the system has to pay in order to get a metric system of space that is consistent across environments.

It is poorly understood how much the geometry of a map in a new environment depends on the previous experience of the rat. It has been shown that in some overtrained animals grid maps emerge very fast in a novel environment[5], but we are far from being able to generalize this observation to all rats, experienced and naïve [35]. It is possible that the perfect grid pattern observed in some rats is a result of the successful generalization of an isotropic distribution of trajectories in 2 dimensions, related to a vast experience in open field environments at the right age of development, possibly starting around P20. Such extensive exposure to the open field is far from being an exclusive natural condition for all rats. It would be interesting to study the adult grid maps of rats raised in a different environment, for example in a system of tunnels. If there is a critical time window when the geometry of the grid system is developed, these animals might have, as adults, grid cells with peculiar properties even after prolonged training in the usual 2 dimensional square boxes. Another interesting aspect of the problem is what happens during this training period. Perhaps extensive learning can compensate for part of the distortion observed in grid maps. The main problem of such a study is that untrained animals exhibit a very poor coverage of 2 dimensional environments, making it hard to compare the geometry of their spatial maps with trained controls.

The network model produces qualitatively similar results over a large portion of parameter space. The delay parameter τ is critical to avoid the collapse of the subpopulation with similar HD into a regime of synchronous dynamics. One wonders whether this delay in synaptic transmission might be realized by NMDA synapses, or
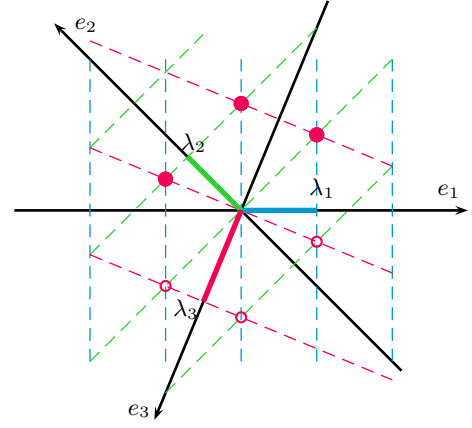

Figure 16: Non-evenly separated grid axes, i.e. not at $\pi/3$ of each other, imply non-even lengths of the axes themselves, to preserve periodicity, resulting in an elliptical arrangement of the vertices. In this example, the grid axes (indicated by the broken lines) are oriented towards at 90, 45 and 157.5 degrees. Orthogonal to them are the three projection axes entering the 2D Fourier transform, at angles e_i of 0, 135, and 247.5 degrees, respectively. The vertices of the grid are arranged on lines at 3 different distances λ_i from each other (and from the origin).

other complex form of coupling among conjunctive cells.

It is possible to extend the network model to incorporate path integration cues into the inputs. Grids in such a network would still be stable even with week spatial inputs, and can serve as spatial codes useful in robot exploration tasks, where the robot has to rely on its own location estimation[44–47].

Acknowledgements

Appendix A: Abstract ellipse model

What does a non-circular, but elliptical arrangement of the grid map imply, geometrically?

If any two of the three grid axes are not separated by 60 degrees, the lengths of the grid axes cannot be equal any more, for the grid to retain its canonical periodicity. As Fig. 16 illustrates, the grid vertices occur periodically along lines (broken in the Figure; those passing through the origin are called grid axes) orthogonal to the three projection axes (the black axes e_i in Fig. 16), provided such lines share 3-way intersections. If the grid is modeled as a simple sum of 3 cosines, the distance between

the broken lines of each orientation is the wavelength $\lambda_i \equiv 2\pi/k_i$ of the corresponding cosine, which has a $\vec{k}$-vector of length k_i oriented at e_i from the x-axis. To preserve grid structure, the inverse wavelength k_3 along the third projection axis e_3 is constrained so that the peaks (red broken lines in Fig. 16) coincide with those along each of the other two projection axes (blue and green broken lines). This leads to two equations relating λ_3 to λ_1 and λ_2 and to the angles between them. In the Figure, it is assumed that $e_1 = 0$, but it is easy to write the following equations considering the general case of nonzero e_1

$$\lambda_3 = \frac{\lambda_1}{\cos(e_2 - \pi/2 - e_1)}\cos(e_3 - e_2 - \pi/2),$$

$$\lambda_3 = \frac{\lambda_2}{\cos(e_2 - \pi/2 - e_1)}\cos(e_3 - e_1 + \pi/2).$$

They lead to a solution for e_3 and λ_3 which may be expressed for example as

$$e_3 = \arccos\left[\frac{\lambda_1 \cos(e_2 - e_1) + \lambda_2}{\sqrt{\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2\cos(e_2 - e_1)}}\right] + \pi + e_1,$$

$$\lambda_3 = \frac{\lambda_1\lambda_2}{\sqrt{\lambda_1^2 + \lambda_2^2 + 2\lambda_1\lambda_2\cos(e_2 - e_1)}}.$$

It is straightforward to see that these relations translate into an equivalent, but simpler relation between the $\vec{k}$-vectors

$$\vec{k}_1 + \vec{k}_2 + \vec{k}_3 = 0. \tag{A1}$$

What determines the exact position of the $\vec{k}$-vectors?

Considering the abstract model discussed in detail in[23], of the development of an individual grid, the model does not distinguish between a would-be pure grid and a conjunctive unit. It assumes that at the end of the developmental process the firing map $\Psi_i(\vec{x})$ of the unit minimizes a functional L, which in its basic version takes the form

$$L = \frac{1}{A}\int_A d\vec{x}[\nabla\Psi(\vec{x})]^2 + \frac{\gamma}{A}\int_A d\vec{x}\Psi(\vec{x})\int_A d\vec{x'}\Psi(\vec{x'})K(|\vec{x'}-\vec{x}|), \tag{A2}$$

where $K(|\Delta\vec{x}|)$ is a kernel expressing neuronal fatigue. The first term of the cost function expresses a penalty for maps that vary too much across space, and a preference for smooth maps. This can be understood from the point of view of the smoothness of the spatial inputs and the smoothness of the neuronal transfer function. The second term of the cost function expresses a penalty for maps in which a neuron has to fire for very long periods of time, subject to neural fatigue. In[23] the grid pattern emerges as a compromise solution between these opposite requirements. There we give two examples of the kernel, which can be treated analytically. The analysis involves going into Fourier space, where the firing map is decomposed into 2D Fourier modes

$$\Psi_i(\vec{x}) = a_0 + \sum_i a_i \cos(\vec{k}_i \cdot \vec{x} + \phi_i) \tag{A3}$$

and the functional becomes

$$L = \frac{1}{2}\sum_i a_i^2 k_i^2 + \gamma_0^2 \tilde{K}(0) + \frac{\gamma}{2}\sum_i a_i^2 \tilde{K}(k_i) \tag{A4}$$

where $\tilde{K}$ simply denotes the 2D Fourier transform of the kernel, and k, again, is the length of the 2D vector $\vec{k}$.

As discussed in[23], among 2D periodic solutions the favored ones are those with a superposition of 3 cosines

$$\Psi(\vec{x}) = 1 + (2/3)\sum_{i=1}^{3}\cos(\vec{k}_i \cdot \vec{x}) \tag{A5}$$

where for simplicity we have omitted the phases by suitably defining the origin.

To model a degree of speed anisotropy, with faster runs e.g. along the axes or along the diagonals of a square environment, one can simply add a quadrupole term to the functional L, the precise form of which is not essential, like the precise form of the kernel K is not essential either. One simple choice is

$$L^{\text{anis}} = \frac{1}{2}\sum_i a_i^2 k_i^2 + \gamma_0^2 \tilde{K}(0) + \frac{\gamma}{2}\sum_i a_i^2 \tilde{K}(k_i)$$
$$+ \zeta \sum_i \left[(k_{xi}^3 + k_{yi}^3)/k_i^3 - 1/\sqrt{2}\right], \tag{A6}$$

where the quadrupole term, with a strength ζ, adds a cost $\zeta(\sqrt{2} - 1)/\sqrt{2}$ for each $\vec{k}$ aligned with the x- or y-axis, and no cost for those oriented at $\pm\pi/4$, hence favoring $\vec{k}$ vectors along the diagonals (rotating the quadrupole by $\pi/4$ one can favor instead those aligned with the axes).

The addition of the extra term makes it hard to get analytical solutions, but numerical minima of the cost function can be found with a gradient descent algorithm that takes into account the additional constraint on the triplet of $\vec{k}_i$ vectors and their amplitudes a_i, $i = 1, 2, 3$. We performed several numerical minimizations of L^{anis} using as initial conditions perfect grids oriented in all possible directions , with the 3 $\vec{k}_i$ at $2\pi/3$ of each other, their length such as to minimize L, and $a_i \equiv 2/3$. The numerical algorithm rapidly converges to a solution with one somewhat shorter $\vec{k}$ vector aligned with one of the two orthogonal directions of faster speed (in Fig.17, $3\pi/4$) and the other two vectors somewhat longer and more than $2\pi/3$ of each other. Whatever the initial orientation of the original symmetric grid, the algorithm finds a solution very close to either of the two "exact" solutions, the one shown in Fig.17 and the equivalent reflection along the vertical or horizontal axis. In this way, when only the single cell level effect is considered, the shorter vector, indicating roughly the direction of the ellipse, stands at $\pm\pi/4$ (or at $0, \pi/2$ if the faster speed are along the axes).
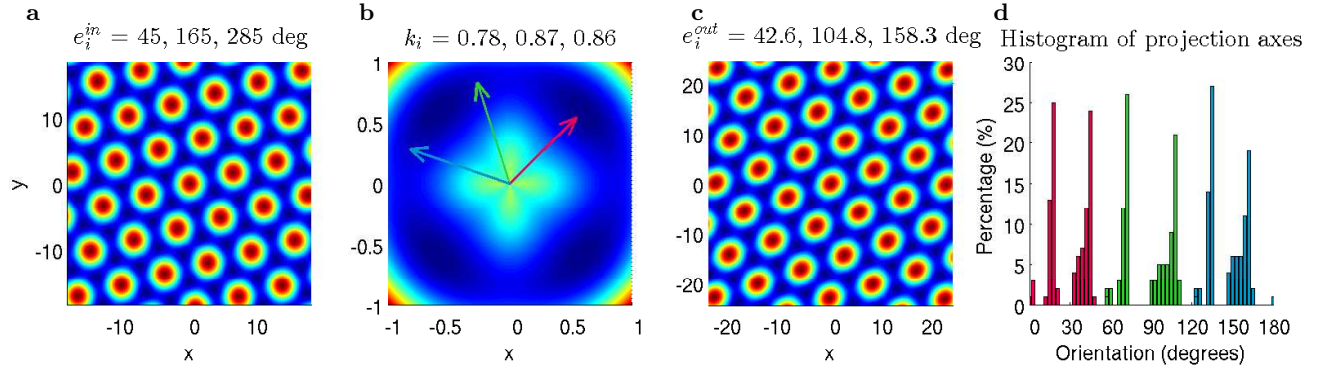
Figure 17: The input (a) and output (c) produced by the gradient descent algorithm that adjusts the $\vec{k}_i$ vectors to minimize a cost function L^{anis} that includes a quadrupole term (panel b, in Fourier space). Depending on the exact initial condition, the algorithm converges towards the theoretically optimal solution (which has angles $e_i = -110, 20, -45$ deg and the two longer $\vec{k}_i$ vectors of identical length) but it stops before reaching it, as solutions with a jitter of ± 2 deg have the same cost, to the fourth significant digit. Panel c shows in different colors the orientation distribution of the three e_is, clustering into two modes. However, the shallow landscape of the cost function (for any reasonable form of the quadrupole anisotropy) indicates that the tight alignment seen experimentally cannot result from single-unit mechanisms alone.

[*] Electronic address: {bailusi,ale}@sissa.it

[†] Electronic address: emilio.kropff@ntnu.no

[1] W. B. Scoville and B. Milner, J. Neurol. Neurosurg. Psychiatry **20**, 11 (1957).

[2] J. O'Keefe and J. Dostrovsky, Brain research **34**, 171 (1971).

[3] J. S. Taube, R. U. Muller, and J. B. Ranck, The Journal of neuroscience **10**, 420 (1990).

[4] M. Fyhn, S. Molden, M. Witter, E. Moser, and M.-B. Moser, Science **305**, 1258 (2004).

[5] T. Hafting, M. Fyhn, S. Molden, M.-B. B. Moser, and E. I. Moser, Nature **436**, 801 (2005).

[6] F. Sargolini, M. Fyhn, T. Hafting, B. L. McNaughton, M. P. Witter, M.-B. Moser, and E. I. Moser, Science **312**, 758 (2006), data available at: http://www.ntnu.no/cbm/gridcell.

[7] L. M. Giocomo, M.-B. Moser, and E. I. Moser, Neuron **71**, 589 (2011).

[8] B. McNaughton, F. Battaglia, O. Jensen, E. Moser, and M.-B. Moser, Nature Reviews Neurosci. **7**, 663 (2006).

[9] M. Fuhs and D. Touretzky, J. Neurosci. **26**, 4266 (2006).

[10] A. Guanella, D. Kiper, and P. Verschure, International journal of neural systems **17**, 231 (2007).

[11] Y. Burak and I. R. Fiete, PLoS Comput Biol **5**, e1000291 (2009).

[12] Z. Navratilova, L. M. Giocomo, J.-M. M. Fellous, M. E. Hasselmo, and B. L. McNaughton, Hippocampus (2011).

[13] K. Zhang, The Journal of neuroscience **16**, 2112 (1996).

[14] A. Samsonovich and B. L. McNaughton, The Journal of neuroscience **17**, 5900 (1997).

[15] D. M. Walters and S. M. Stringer, Biological Cybernetics **103**, 21 (2010).

[16] N. Burgess, C. Barry, and J. O'Keefe, Hippocampus **17(9)**, 901 (2007).

[17] L. M. Giocomo, E. A. Zilli, E. Fransén, and M. E. Hasselmo, Science **315**, 1719 (2007).

[18] N. Burgess, Hippocampus **18**, 1157 (2008).

[19] M. E. Hasselmo, Hippocampus **18**, 1213 (2008).

[20] E. A. Zilli and M. E. Hasselmo, The Journal of neuroscience **30**, 13850 (2010).

[21] J. Koenig, A. N. Linder, J. K. Leutgeb, and S. Leutgeb, Science **332**, 592 (2011).

[22] M. P. Brandon, A. R. Bogaard, C. P. Libby, M. A. Connerney, K. Gupta, and M. E. Hasselmo, Science **332**, 595 (2011).

[23] E. Kropff and A. Treves, Hippocampus **18**, 1256 (2008).

[24] D. L. Garden, P. D. Dodson, C. O'Donnell, M. D. White, and M. F. Nolan, Neuron **60**, 875 (2008).

[25] R. F. Langston, J. A. Ainge, J. J. Couey, C. B. Canto, T. L. Bjerknes, M. P. Witter, E. I. Moser, and M.-B. B. Moser, Science **328**, 1576 (2010).

[26] T. J. Wills, F. Cacucci, N. Burgess, and J. O'Keefe, Science **328**, 1573 (2010).

[27] H. Stensland, T. Kirkesola, E. Moser, and M.-B. Moser, Society for Neuroscience abstract **101.14** (2010).

[28] L. L. Colgin, E. I. Moser, and M.-B. Moser, Trends in Neurosciences **31**, 469 (2008).

[29] M. Fyhn, T. Hafting, A. Treves, M.-B. Moser, and E. Moser, Nature **446**, 190 (2007).

[30] C. N. Boccara, F. Sargolini, V. H. H. Thoresen, T. Solstad, M. P. Witter, E. I. Moser, and M.-B. B. Moser, Nature neuroscience **13**, 987 (2010).

[31] A. Dhillon and R. S. G. Jones, Neuroscience **99**, 413 (2000).

[32] M. S. Petkovic, *Famous Puzzles of Great Mathematicians* (American Mathematical Society, College Station, Texas, 2009).

[33] C. Barry, R. Hayman, N. Burgess, and K. Jeffery, Nat. Neurosci. **10**, 682 (2007).

[34] F. Savelli, D. Yoganarasimha, and J. Knierim, Hippocampus **18(12)**, 1270 (2008).

[35] C. Barry, J. O'Keefe, and N. Burgess, Society for Neuroscience abstract **101.24** (2009).

[36] J. H. Zar, *Biostatistical Analysis (4th Edition)* (Prentice

Hall, 1998).

[37] T. Solstad, E. Moser, and G. Einevoll, Hippocampus **16**, 1026 (2006).

[38] E. Rolls, S. Stringer, and T. Elliot, Network: Comput Neural Sys **15**, 447 (2006).

[39] I. R. Fiete, Y. Burak, and T. Brookings, The Journal of neuroscience **28**, 6858 (2008).

[40] B. Si and A. Treves, Cognitive Neurodynamics **3**, 177 (2009).

[41] T. van Haeften, L. Baks-Te-Bulte, P. H. Goede, F. G. Wouterlood, and M. P. Witter, Hippocampus **13**, 943 (2003).

[42] A. Treves, Journal of Computational Neuroscience **14**, 271 (2003).

[43] D. Derdikman, J. R. Whitlock, A. Tsao, M. Fyhn, T. Hafting, M.-B. B. Moser, and E. I. Moser, Nature neuroscience **12**, 1325 (2009).

[44] B. Si, J. M. Herrmann, and K. Pawelzik, in *Proceedings of the International Conference on Natural Computation* (2007), pp. 177–182.

[45] M. Franzius, R. Vollgraf, and L. Wiskott, J. Comput. Neurosci. **22**, 297 (2007b).

[46] D. Samu, P. Erős, B. Ujfalussy, and T. Kiss, Biological Cybernetics **101**, 19 (2009).

[47] M. J. Milford, J. Wiles, and G. F. Wyeth, PLoS Comput Biol **6**, e1000995 (2010).